

IMPLICATIONS OF EU EVALUATION PRACTICES: FROM A POSITIVIST APPROACH TO A REALIST ONE IN EVALUATING 2007-2013 COHESION POLICY

Ioana Spanache

Complutense University of Madrid

Madrid, Spain

ioana.borcan09@gmil.com

Abstract

The present paper aims to analyse the recommended evaluation practices at EU level and their possible implications on the results and recommendations formulated at the end of the evaluation process. To do so, I have built a model for analysing the theoretical assumptions behind the evaluation practices, and applied it to reports and TORs of evaluations requested by the European Commission regarding the Cohesion Policy in the 2007-2013 financial perspective.

Here, one might ask: Why is it relevant to know and analyse the theoretical assumptions behind evaluation practices in general and those recommended at European Union level in particular? By including certain criteria in the terms of reference, different interest groups (in this case the European Commission, representatives of national, regional and local institutions, non-governmental organizations, representatives of the private environment, etc.) decide what can be considered as valuable or not.

Compared with what authors Hoerner and Stephenson (2012) have found regarding the evaluation practices that covered the Cohesion Policy in the 2000-2006 financial perspectives – a predominance of the positivist paradigm in evaluation practices, this article concludes that the dominant paradigm between 2007 and 2013 is that of realist evaluation. Both the analysis model and the

results obtained after applying to eight Cohesion Policy work packages are detailed in this article.

Keywords

Cohesion policy, evaluation practice, realist evaluation, scope of evaluation

1.ELEMENTS OF A NEW MODEL OF ANALYSIS

For the construction of the model of analysis I chose some of the descriptors previously used by authors Hoerner and Stephenson (2011) in order to analyse the theoretical approaches of the evaluation practices that the European Commission recommended and applied in the evaluation of the Cohesion Policy. To these five descriptors, I have decided to add other five descriptions so as to make a more comprehensive and deeper analysis. I will further detail the significance of each descriptor included in the model of analysis and in the context of three evaluation paradigms – pozitivism, realism and constructivism. Moreover, I will also specify the reasons behind the proposed new descriptions and explain their relevance for the built model.

Thus, the five characteristics used by the abovementioned authors are: epistemological assumptions; methods of evaluation / research and data collection; type of approach - top-down or bottom-up / micro- or macro-level analysis; engagement of interest groups; the role of the evaluator. To these, I chose to add the following: ontology; the connection between cause and effect; attitude towards values; the purpose of the evaluation; and the type of learning. Further, I look at each characteristic of how practical approaches relate to the theoretical perspectives in the field of evaluation.

1.1.Ontology and Epistemology

When they speak about the epistemological assumptions behind evaluation practices, Hoerner and Stephenson refer to the evaluator's philosophical approach towards the process of acquiring new knowledge and evaluation as a social-scientific endeavor. Here, the two give arguments that epistemological assumptions can determine whether evaluators aim to produce objective, generalizable and of universal value results, or if they admit the existence of different influencing factors such as the unique way in which each individual perceives the world, which does not allow making generalizations. (Hoerner and Stephenson 2011, 702)

Unlike the two authors, I chose to also look at the ontological assumptions behind the evaluation practices recommended by the European Commission and applied by the evaluators, in order to have a deeper analysis of both aspects relating to the nature of knowledge and how it is perceived by evaluators or TOR authors, as well as aspects related to the nature of reality and how it is addressed by the two types of actors.

On the other hand, the two descriptors are used separately in analysis also by other authors, such as Guba and Lincoln (Hoerner and Stephenson 1989, 84), who compare positivist and constructivist paradigms from three perspectives: ontology, epistemology and methodology.

Thus, when it comes to the ontological specificity of an evaluation practice, the two consider whether reality is perceived as being unique and independent from the interests of its observer and determined by certain natural laws often characterized by cause and effect relationships (in the positivist paradigm) or as something that can take up many different forms, due to the existence of multiple and socially constructed realities. And, in the latter perspective, the "truth" is considered to be built, similar to the case of the constructivist paradigm. Therefore, the ontological assumption behind certain evaluation practices will refer to the perception of reality and truth. And for analysis I will use the following questions: What is the nature of reality?; Can it be known?; How is the reality perceived by evaluators / authors of terms of reference? -

objective, subjective, complex, characterized by immutable natural laws, built through socialization, determined by the context, etc.

Epistemologically, Guba and Lincoln make the difference between an objective dualist approach (in the case of positivism), in which the researcher can characterize the studied phenomenon faithfully, while remaining detached from it (a dual subject-object relationship) and abstracted from any considerations regarding value; and a subjective monist approach (in constructivism), in which the researcher and the object of research are interconnected in such a way that the results of the research are literally the creation of the research process (1989, p. 84). Thus, in the first situation, the results of the research are perceived as objective and generalizable, while in the second situation it is considered that they cannot be generalized, and that they may differ from case to case depending on context and factors of influence. The questions that I will use to analyse the epistemological perspective within the TOR and the evaluation reports are: What is the nature of knowledge? Can reality be known?; Can we reach the truth through research?; Can the results of the evaluation be considered as objective?; Is there an absolute truth about them?

Here, it should be noted that the approach used in the present paper regarding the concepts of ontology and epistemology is specific to authors in the field of public policy and program evaluation. In the paper, I use the two concepts as they are understood by these authors, and not according to the approach used in philosophy.

1.2. Methodology

For the analysis of this component, I have looked at the extent to which the evaluations requested (through ToRs) and conducted in the case of EU's Cohesion Policy include the use of quantitative and/ or qualitative methods of data analysis and collection. Here, I suggest the following types of questions: Are qualitative or quantitative methods used?; Is there a predominant evaluation method that is being used?; How are data collected? (through quantitative or qualitative techniques and instruments).

On the other hand, the type of methods and techniques used in an evaluation can provide us with information regarding both epistemological and ontological assumptions underlying an evaluation, but also the extent to which interest groups are involved, the importance attributed to the input offered by each category, the type of approach (top-down or bottom-up), about attitude towards causality, attitude towards values, and even about the purpose of the evaluation. That is why, in most cases, I have started the analysis of the TOR and the final evaluation reports by first identifying and reflecting on used evaluation methods and data collection techniques and tools.

If we look again at the comparative analysis done by Guba and Lincoln (1989, 84), we will see that an "interventionist" methodology in which the factors of influence are completely excluded in order to ensure that the conducted research can explain nature exactly as it is and how it works, which leads us to think about predictability and control, specific to a positivist paradigm. On the other hand, a hermeneutical methodology involving a continuous dialectic process of analysis, critique, reiteration, and reanalysis, through which a common case is constructed, rather indicates a constructivist paradigm in the evaluation. Meanwhile, Bachtler and Wren believe that the methodological diversity specific to the evaluations of Cohesion Policy programs can also be seen as a strong point. (Bachtler and Wren 2007, 114) And this is due to its ability to adapt according to the context and to the specificity of the evaluand. The two put this diversity on account of the differences in the type of data collected, the degree of sophistication of the application and the types of questions addressed. Thus, Bachtler and Wren distinguished between bottom-up evaluation methodologies based on sociological surveys addressed to beneficiaries and top-down methodologies based on input-output impact evaluation models and process studies on how Structural Funds were implemented. (Bachtler and Wren 2007, 114)

Following the analysis of the three evaluation reports, Hoerner and Stephenson conclude that the positivist approach prevails in the evaluation of cohesion policy specific programs (Hoerner and Stephenson 2011, 714). However, as I will show in my analysis in this paper, I believe that the chosen methodology and the type of paradigm that guides the evaluation practices depend to a great extent on

the purpose pursued through evaluation (what is to be achieved by doing it). Moreover, I believe that all of these are also influenced by the objectives pursued by the evaluated programs.

If these objectives are predominantly of a quantitative nature - number of km built or rehabilitated, number of jobs created, etc., the used methodology will also be rather quantitative, specific to a positivist approach.

On the other hand, another argument to support the importance of methodological aspects in the practice and the theoretical approach to evaluation is due to their characterization by Alkin and Christie as one of the branches of evaluation (in Alkin 2013, 12). The two authors use the metaphor of a tree with roots and main branches and secondary branches for the organization of evaluation theorists around certain criteria set by them. Thus, along with the use of evaluation and valorization, methods and methodology are considered as one of the main criteria/ branches of the theoretical approaches in evaluation. The latter is placed by the two right in the middle of the tree, being an illustration of evaluation guided especially by research methodology. And the authors included in the category of this branch are those who are most concerned with obtaining the most rigorous knowledge, despite the existing contextual constraints.

1.3.Causality

In this case, we can pursue the extent to which evaluators and Tor authors consider the causality relationships within the evaluated program as being important, as same as the attitude of the two categories of actors towards them. Through this, I intend to find out if the links between cause and effect are seen as something clear and that can be known, or if they are addressed through an increased attention given to the specific perspectives of different interest groups. The question here is: "Is it expected for the evaluator to be able to identify the causal relationships within the evaluated program?"

If, in the positivist perspective, it is considered that the relationships between causes and effects can be determined and analysed through scientific research, in

post-positivism, a certain degree of uncertainty regarding the results is accepted (Christie and Alkin 2013, 17). Similar to the position in the post-positivist paradigm is that specific to the pragmatist paradigm. However, unlike all these, in constructivism, causality relations are seen as impossible to identify because of the interdependence existing between them, as the elements under analysis are mutually influencing each other at all times. (Christie and Alkin 2013, 17)

Huey-Tsyh Chen (1990, 1992, 2005) is one author that has dedicated a lot of its work to theory-based evaluation. On the other hand, other theorists that have brought substantial contributions to it are Carol Weiss (1997) and Freeman Rossi (1992). And this type of evaluation was later taken over and adapted by the European Commission in the form of theory-based impact assessment. Alongside to counterfactual impact evaluation, TBIE is one of the recommended evaluation approaches by the Commission for the 2014-2020 programming period.

1.4. Involvement of interest groups

This component is also found in the model of analysis developed by Hoerner and Stephenson (2011), the two distinguishing between evaluations that aim only at the participation of those interest groups directly involved in the implementation of the evaluated program and those aiming to ensure a more comprehensive and inclusive participation from the community, no matter how marginal the influence of the program on the groups concerned (702-703). In this regard, the two authors have grounded their fourth feature of analysis on the constructivist approach of evaluation from Guba and Lincoln (2001).

Here, Hoerner and Stephenson describe the minimum level of stakeholder engagement as being a situation specific to the positivist paradigm, in which the interested groups are approached only passively as economic actors and as an object of analysis (2011, 706). In my opinion, in this situation, the evaluation would only include the clients (the actors who have requested the evaluation) as active participants in the evaluation process, leaving behind the direct beneficiaries, other actors involved in program implementation, as same as other

interested parties such as representatives of academia or non-governmental organizations, who would have no role to play. On the opposite side, there is the maximum level of involvement of interest groups, where the latter would be also used in building the evaluation design. In this case, the evaluation process would be similar to that suggested by the constructivist and participatory approaches, in which the evaluator tries to co-opt all affected parties through the evaluated program, including even the more reticent actors.

Regarding the relationship of this descriptor with other elements of the model of analysis, we can say that stakeholder engagement can also provide us with information on the type of approach and mode of analysis (top-down / bottom-up approach, macro / micro analysis), about the epistemology and ontology specific to an evaluation, the purpose of the evaluation, and in particular, the attitude of evaluators/ authors of ToR towards values. An evaluation that involves the use of questionnaires and the organization of interviews with the direct beneficiaries of the intervention may also imply a higher degree of involvement of interest groups, as same as elements specific to the bottom-up approach and micro-analysis. On the other hand, an evaluation that involves the engagement of direct beneficiaries, and their participation in building its design, may indicate an ontological and epistemological perspective specific to constructivism. Regarding the relationship between stakeholder engagement and the other two characteristics, attitude towards values and purpose of evaluation, I will detail below.

1.5. Attitude towards values

In this case, Michael Scriven (1967, 1972a) and Robert Stake (1967, 1974) are considered to be the reference authors who introduced the valuation concept. Moreover, Scriven is named as the author who claimed for the first time that evaluation would not be evaluation without values. (Christie and Alkin 2013, 31) Thus, with regard to this characteristic, we can pursue several dimensions of analysis regarding evaluation practices. On the one hand, we have evaluations that do not acknowledge the existence and role of values in their process and, on

the other hand, are the evaluations that take into account the existence and influence of values, but the latter can also be divided into two other categories as follows: 1. Evaluation that takes into account the existence of values, but in which the evaluator attempts to reduce and even eliminate their influence through the chosen methodology and the control of variables; 2. and evaluations in which values are integrated by the evaluator in the evaluation process, where they are considered to be part of the program and evaluation context itself, and part of the different views about the world and about the program held by interest groups.

Therefore, the questions that will be applied in the context of this descriptor are: What attitude does the evaluator/ ToR author have towards values?; Do they recognize their influence on the program and/ or on the context?; Do they try to integrate them into the evaluation process? And if so, to what extent? The attitude towards values is closely related to the involvement of interest groups. Integrating stakeholder values requires the use of participatory evaluation methods.

1.6.Type of approach – top-bottom or bottom-up (macro/ micro analysis)

The descriptor regarding the type of approach is also found in the work of authors Hoerner and Stephenson (2011). And when it comes to it, the two refer to the way in which data are collected for undergoing an evaluation, either from secondary sources, from the national and / or regional public administration and from statistical agents, or collected directly from the stakeholder groups involved, in order to capture local experiences and opinions (Hoerner and Stephenson 2011, 702). Thus, in close relation to the type of approach, there are the descriptors of interest groups engagement, methodology, but also attitude towards values. By analysing the three we get an overview on the nature of the evaluation direction.

And when it comes to the two types of approaches, Armstrong and Wells provide the following examples of used methods (2006, 264): statistical methods such as regression analysis, econometric models, input-output analysis – specific

to a top-down approach with macro-analysis; and the sociological survey conducted at the level of program beneficiaries - in case of the bottom-up approach with micro-level analysis. On the other hand, the two authors draw attention to the fact that qualitative methods of evaluation and research can also be successfully used in the positivist paradigm. In this regard, they give as example the case of Great Britain and the long tradition of combining qualitative and quantitative methods in monitoring and evaluating the cohesion policy (King 1990, Armstrong and Wells 2006, 264). Here, the difference between the positivist paradigm and another paradigm, like realism or constructivism, is given by the way in which the collected data is analysed - through statistical methods (in the case of the positivist paradigm) or qualitative methods (in the case of other paradigms). In fact, such an analysis would in fact be specific in terms of the used methodology to the post-positivist paradigm, which also supports qualitative methods.

One of the reasons why Hoerner and Stephenson concluded that the preferred paradigm in the evaluations carried out on the 2000-2006 Cohesion Policy is positivist, is due to the fact that the analysis specific to this set of evaluations involved rather a top-down approach, and was undertaken at macro level. Through this paper, I seek to find out what type of approach is used predominantly in the evaluations for the Cohesion Policy program from the 2007-2013 financial perspective and whether the conclusions of the two authors are also confirmed in the latter case.

The questions which I will use to analyse the current characteristic are: What kind of analyses are conducted - at a meta level or more in-depth studies undertaken at a smaller scale?; Is there a pre-established set of questions set at a central level by evaluators / commissioners? Or are the questions established at local level?; Can the evaluation results be considered generalizable? Or does the evaluation recognize the importance of context and socio-economic determinants?; What kind of logic prevails in the process of evaluation - deductive or inductive?.

1.7. The evaluator's role

To draw conclusions about the role of evaluator from the terms of reference and evaluation reports I will use the following types of questions: How are evaluators conceptualized? Are they perceived as neutral or rather as an independent factor in the evaluation process? Here, the two extremes between which the descriptor regarding the role of the evaluator are: the evaluator as a disinterested actor, in the positivist paradigm; and the evaluator as a "passionate participant", actively involved in facilitating a "multi-voice" reconstruction of the constructions belonging to him/her and to other participants, as orchestrator and facilitator of the research process, responsible for ensuring representation of all stakeholders affected by the program. (Guba and Lincoln 1994, 115)

In addition to analysing stakeholders' involvement and how the evaluator's perspective impacts evaluation reports, I will seek to find out by applying the model of analysis, what kind of roles of the evaluator can be found in the documents reviewed: as an actor responsible for the transformation of society, as a neutral researcher, an educator or "merchant power" (Skolits, Morrow and Burr 2009). In doing this, I refer to how the documents describe the interaction of evaluators with interest groups, as same as the type of methods used, the scope of evaluation and the easiness or difficulty with which we can infer from the analysed documents data about the evaluator's perspective.

1.8. The scope of evaluation

Regarding the link between the scope of evaluation and the theoretical approaches from the field there are much to say, but I will start by mentioning the research questions which I will use in the analysis of this descriptor/characteristic: When is the evaluation conducted? - Ex-ante, mid-term or ex-post?; To which elements of the process of policy/program-making is given more importance according to the scope of evaluation? (Results, objectives, implementation process); To what extent is evaluation used as a facilitator for

learning?; What is the final goal of the evaluation?; For what purposes are going to be used its results?.

The scope of evaluation is a subject that was approached by several academics and practitioners in the field. Some names in this regard are: Michael Quinn Patton, Jennifer C. Greene, Eleanor Chelimsky, Lee Cronbach, Joseph S. Wholey, Donna Mertens, Patricia J. Rogers, Peter Dahler-Larsen, and others. Thus, evaluation was analysed in regard to its role in ensuring financial responsibility, in providing justification and legitimization for the implemented public policies and programs, in transforming society, for organizational learning and performance improvement of an institution, in generating new knowledge, and even in the context of improving democratic practices, so as to make the world a better place. My hypothesis in this regard is that, when an author makes his preferences known on a certain type of methodological, ontological and/ or epistemological approach, he/she implicitly expresses preferences for certain purposes of evaluation.

Thus, Eleanor Chelimsky (2006) distinguishes between the original purpose of evaluation on ensuring financial accountability, and other purposes that are derived from the former, like deepening knowledge in a particular field, strengthening administrative capacity and improving performance in organizations, as same as informing citizens. And all this ultimately contribute to improving the functioning of democratic processes in a society (33-56). Also, the view of author Jennifer Greene is similar to the one illustrated by Eleanor Chelimsky. Greene (2013, 208- 218) believes that evaluation can help make the world a better place and that it can be used in promoting a more inclusive approach towards interest groups, by integrating cultural and contextual aspects in its process. The same author has even built a model containing elements necessary for designing evaluation theories based on democratic principles (Greene 2006, 135). Moreover, Greene also offers a good example of the connection existing between the scope of evaluation and the epistemological assumptions underlying it, when she says that her model of evaluation based on democratic principles should include a constructivist and interpretivist framework.

On the other hand, in 1996, Michael Patton identified a "menu" of evaluation purposes developed in response to an article written by Michael Scriven (1991), in which the latter has supported the existence of a formative versus summative dichotomy. Thus, Patton (142) argues that this dichotomy was useful during the development of the evaluation field and that nowadays we should rather accept the existence of multiple purposes of evaluation and of a diversity of values. The author believes that evaluation serves also other purposes than those appointed by Scriven, such as the generation of knowledge about the effectiveness of programs; the development of programs and organizations; focusing management efforts; creating learning organizations; empowering participants; directly supporting and improving interventions (by integrating evaluation in those interventions); and stimulating critic reflection in order to improve evaluation practices.

The opposite of these approaches is represented by the CIPP model built by Daniel L. Stufflebeam (2013, 243), that aims at conducting evaluations oriented on decision-making and on ensuring financial accountability. Stufflebeam (2013, 243- 258) offers a strong role to aspects of context and core values identified in the object of evaluation, but he describes the epistemological approach of its model as being rather objectivist and not relativist.

1.9. The type of learning

In 2015, the evaluator Richard Hummelbrunner (2015) managed to create an exploratory innovative framework designed to contribute to improving consistency in evaluation. For this, the author has combined several approaches for a more comprehensive evaluation processes. His model can be used by both evaluators and commissioners before drawing up the terms of reference or designing an evaluation, so that evaluation can be better tailored in regard to its aims. Hummelbrunner notes that, while usually these have been used separately, the three typologies can bring more added value when used together, due to their complementarity.

The conceptual framework constructed by the author is based on three dimensions of analysis and the typologies affiliated: learning in evaluation; concepts from systems theory; and values in evaluation. To illustrate the first dimension, Hummelbrunner starts from Gregory Bateson's statements about how knowledge is transmitted within social systems (Bateson 1972 cited. Hummelbrunner 2015, 17). In his opinion, knowledge is not "transferred" in a passive way from the transmitter to the receiver but it is rather gained through an active process. In this process, it is also relevant the way in which individuals select and retain information based on the transmitted signals. Thus, the receiver does more than simply take the signals as such, and he/she assigns meanings to them before transforming them into information. In other words, the resulting information consists of the initially received signal plus the meaning assigned to it by the receiver after passing it through the filter of their own conceptions and values.

Making a parenthesis, here we can draw conclusions about the importance of values including the evaluation. And if we try to integrate the claims of Gregory Bateson in one of theoretical approaches of evaluation used in this paper, we may say that the author in question seems to have a constructivist perspective. If individuals "build" information based on the signal received from the transmitter and the meaning they attribute to that signal, then it can be considered that reality, which consists of the sum of information produced by individuals, is in turn a construct that can be as subjective as the meanings assigned to the received signals by receptors.

Returning to Bateson, learning represents also a change process towards a previous situation that can take place on multiple levels, depending on how it deals with the relationship between trial and error (Bateson 1972 cit. Hummelbrunner 2015, 17-18):

1. Single-loop learning involves correcting errors in a pre-established context or set of alternatives. The latter do not change;
2. Double-loop learning involves doing a change in type 1 learning, which can mean either to produce a change in the set of alternatives on which choice is made, or a change in how the sequence of experience is punctuated. This type of learning implies that the activity takes place in a changing context, and it is

therefore important to recognize the pattern according to which there are changes in the context, as same as the implications of assigning meanings have on an act or on behavior;

3. Triple-loop learning is the next step. As type 2 learning refers to type 1, type 3 refers to type 2, and it involves reflection on and a change in type 2 learning.

Hummelbrunner goes even further in deepening the three types of learning described by Bateson and inspiration from other authors who have extended the model in question and have it applied in organizational development and management - Argyris and Schön (1978), Fiol and Lyles (1985), Senge (1990), Miner and Medes (1996), Arthur and Aisman-Smith (2001). Thus, in a single loop learning becomes his specific question "We do things right?" The double-loop learning is equivalent to the question "We do the right things?" Triple loop learning and corresponds to the following question - "How do we determine to be?".

For this element of the model of analysis, I followed the link between the predominant evaluation approach and the types of learning as conceptualized by author Richard Hummelbrunner (2015). By providing answers to questions - Which elements of a program / policy are being given attention during the evaluation process?; Is the evaluation design suited to a complex/ dynamic situation?; Does evaluation include references to the learning mechanisms behind an intervention?; I tried to find out to which type of learning are the analysed evaluations more close to: single loop learning, double loop learning (which also includes the single loop), triple loop learning (that includes the other two).

The model of analysis from Annex no.1 represents a synthetic view of those described above and it shows how the ten characteristics are reflected in the different theoretical approaches specific to the field of evaluation. It is based on five characteristics used by Hoerner and Stephenson (2011), to which I have added the other five characteristics I have proposed. Moreover, I have also tried to bring certain changes to the five dimensions of analysis suggested by the two authors in 2011, by presenting more broadly the implications that they might have when it comes to theoretical perspectives and by applying them together

the new proposed five descriptors to terms of reference and evaluation reports for the 2007-2013 cohesion policy.

2. ANALYSIS OF EVALUATION PRACTICES IN THE 2007-2013 FINANCIAL PERSPECTIVE – APPLYING THE MODEL

2.1. Analysis of Work Package 0: Data collection and quality assessment

Table no. 1: Analysis of Work Package 0: Data collection and quality assessment

No.	Descriptors	ToR	Final report of evaluation
1.	Ontology	Reality can be known through research	Reality is considered objective
2.	Epistemology	Going through the terms of reference document, I cannot say that I have encountered aspects that indicate the possibility of judging the results as lacking truth. Errors may occur, but because of mistakes made in the research / evaluation, mistakes that can be corrected on the other hand. Evaluation results are regarded as being objective. This is due to the fact that the established indicators are	The mechanisms behind definitions aren't analysed. Evaluators take into account only the perspectives of the following actors: European Commission, Member States (through AIRs), and Managing Authorities; which indicated a higher importance given to actors from the central level. Only quantitative data matters. And the programs are considered as working or not, there

		<p>quantitative, as well as the methodology, which is described in more detail.</p> <p>About interviews and their contribution to the evaluation process, the authors mention only few details. Here, it is possible that the research design might remain in the task of evaluators.</p>	are no nuances in this regard.
3.	Methodology	<p>Analysis of quantitative data and analysis and collection of qualitative data (interviews)</p> <p>The basic indicators, as same as the future ones are quantitative.</p> <p>Example: number of persons, number of km, number of projects, etc.</p>	Mixed methodology of data collection and quantitative methodology of analysis
4.	Cause and effect	The TOR do not include an analysis of causality relations.	The same as in the case of the ToR.
5.	Attitude towards values	Values are not mentioned.	The same as in the case of the ToR.
6.	Type of approach: top-down, bottom-up	Evaluation performed widely in all EU Member States and covering a large number of NUTS level 3 on their territory	Top-down approach and an analysis undertaken at macro level

		<p>There is a default set of assessment questions established at central level by the authors of TOR</p> <p>The evaluation results can be considered generalizable deductive logic</p> <p>Data collected at local and national level are then collected and analysed at central / European level</p> <p>Indicators and methodology are set at central/ European level, and the mechanisms for data collection and measurement of results are the responsibility of Member States and institutions at central or regional level, where applicable.</p> <p>Afterwards, the data are collected from all Member States and analysed at central / European level.</p>	
7.	Stakeholder's involvement	It is suggested the involvement of the following stakeholders:	Evaluators interview representatives from the Managing Authorities,

		<ul style="list-style-type: none"> - Evaluation Unit, DG Regio - the winning team will have to attend a meeting at the beginning of the evaluation and other three progress meetings with representatives of the Unit at the headquarters of the latter; - the European Commission represented by a steering group made up of members of several Directorates from DG Regional Policy and Urban Development and other DGs concerned; - Member States - through representatives of the MAs and other relevant institutions to be interviewed when this is deemed necessary. 	but not for learning more about their perspectives, but for obtaining data of quantitative nature.
8.	Evaluator's role	The evaluator has the freedom to choose other methods of evaluation, in addition to those already proposed in the call for evaluation.	Neutral and independent evaluator.
9.	The purpose of evaluation and attitude towards	Ex-post evaluation - summative Purpose of evaluation =	Increased focus on results, financial accountability and

	results, objectives, implementation process, the role of assessment in learning	financial accountability when it comes to used resources; decision making; legitimization in front of taxpayers Increased focus on results and performance	justification of programs; implementation (just for defining the indicators)
10.	Type of learning	Single-loop learning	Single-loop learning

Analysis of TOR

In the section dedicated to the purpose and context of evaluation, the terms of reference (8) state that the purpose of the evaluation at case is to gather, evaluate and revise terms of quality, where necessary, the data reported by the Managing Authorities of the Member States. Also, we can observe from the text the authors' wish to put more emphasis on financial responsibility when it comes to the resources used. TOR authors believe that this is due to the debate stimulated by the publication in April 2010 of ex-post evaluation results for cohesion policy in the 2000-2006 financial perspective.

Moreover, it is noted that the same provisions will also underpin the debate on the evaluation and implementation of cohesion policy for 2014-2020, the Structural Funds reform being focused on results and performance.

Regarding the use of evaluation results is noted that they will be useful to Member States and MAs in developing monitoring systems related to the 2014-2020 financial perspective. And those systems are even more important due to the fact that the allocation of resources in that interval will partially depend on the supply / reporting of reliable data about the performance of programs. (TOR ex-post from 2007 to 2013, 9)

More specifically, the evaluators who would win the contract for this assessment were to collect and assess in terms of quality the data reported by MAs in annual implementation reports in 2012 and 2013, and to assess the feasibility of collecting financial data broken down by theme and priority for the regional NUTS levels 2 and 3.

Regarding the use of evaluation results is noted that they will be useful to Member States and MAs in developing monitoring systems relating to the 2014-2020 financial perspective. And those systems are even more important as, provision of resources in the interval will depend partly on the supply / reporting of reliable data about the performance of programs. (TOR ex-post from 2007 to 2013, 9)

More specifically, the evaluators who would win the contract for this assessment were to collect and assess in terms of quality data reported by MAs in annual implementation reports in 2012 and 2013, and to assess the feasibility of collecting data financial broken down by theme and priority for the regional NUTS levels 2 and 3.

When it comes to methodology, the TOR suggest to evaluators to use the following methods and evaluation techniques (TOR ex-post from 2007 to 2013, p.14): desk research, which involves an analysis of the annual implementation reports and of other useful documents written by Member States; analysis of data stored in databases of DG for Regional and Urban Development (SFC2007 and Infoview) - in this case, the evaluation team will receive access; interviews with representatives from Member States, MAs and Intermediate Bodies (IBs), where deemed necessary; and other methodological approaches, as appropriate, to be specified by the evaluation teams participating in the call selection.

By reading the deliverables section, we can see that the evaluators are required to analyse to which extent have the annual implementation reports have improved qualitatively, between 2013 and 2012. Therefore, here, we can consider that it is the case of some sort of metaevaluation.

Towards the end of the reference document, it is stated that, after the completion of the evaluation report, the latter will be metaevaluated by certain services of the European Commission on the basis of quality criteria specified in the Guide for Evaluation of Socio Economic Development.

For the analysis of certain indicators, the evaluation team would use methodologies for desk research and the interview method for interviewing several representatives of the MAs and the Member States in general. Here, the authors of TOR warn the evaluators regarding the different practices existing in Member States, a fact that requires for them to contact several institutions in the

same Member State (eg in each region, where administrative systems are decentralized). Also, in this case, we can see that all indicators are quantitative - expressed generally as number of people, number of km, project number, etc.

No.	Evaluation questions proposed in TOR	Observations
1.	Is the definition of the indicator and the methodology to be used clearly documented? (Give an indication of what the definition is and what guidance is provided on how the data should be collected and aggregated, or in countries where there are a large number of programmes, what typically is the case.)	Here, a new indicator could be added - compliance in methodology
2.	Is the same definition and methodology used in each operational programme, i.e. by each managing authority? If not, how do they differ? Is any attempt made to ensure consistency of the definition of the indicator and the data reported within countries?	The coherence between the definition given to the indicator and the data reported in the Member States
3.	Do the data reported relate to actual outcomes in the case of all programmes as opposed to expected or planned outcomes? Is the basis of the data reported clear in all cases?	Comparative analysis between results obtained and results expected - element specific to the positivist approach in evaluation
4.	Is there any quality control mechanism that checks the data	New indicators: data accuracy and plausibility

	before it is submitted? If yes, where it is located (within or outside the Managing Authority, central or regional government)? Are any plausibility checks carried out on the data within the Managing Authority?	
5.	Does the central government aggregate the data across programmes to calculate a national figure for each indicator?	This question reflects a quantitative methodology
6.	What are the strengths and weaknesses of the reporting for each MS?	This question involves a more in-depth analysis compared to the quantitative indicators

Summarizing the comments made above regarding the evaluation questions suggested by the authors of the terms of reference, we can say that the issues covered by them are mostly on the accuracy of data provided by Member States for the indicators listed above. Furthermore, I wish to draw attention to question number three, that aims at doing a comparative analysis between the results obtained in the Member States and the expected results. Such analysis is typically specific to the positivist approach in evaluation.

Analysis of the final report of evaluation

In the abstract from the beginning of the evaluation report it is noted that the guide documents provided by DG Regio for reporting on the financial perspective 2007-2013 have introduced new core indicators to facilitate aggregation and comparison of data from implementing similar programs in Member States. It also discusses the introduction of new indicators for 2014-2020. The approach developed by the team that won the contract for the ex-post evaluation of the programs implemented under the Cohesion Policy in 2007-2013 range was to assess the physical and the financial data reported by MAs in annual implementation reports (AIR) for more than 300 Operational Programmes (OP) funded through ERDF and CF. The latter led to the following:

identification of problem areas for expenditure incurred within ERDF and CF NUTS3 regions; review the annual implementation reports for 2012 and 2013, which were targeted approach by both key indicators and the future, and other short-term results and long-significant; analysis of reporting on major projects described in the RAL 2013; draw conclusions on best practices recorded in monitoring programs to reduce greenhouse gas emissions. Thus, the introduction to the final evaluation report seems to contain exactly all 4 tasks included in the terms of reference.

2.2. Analysis of Work Package no. 2 – Support for SMEs - Increasing research and innovation in SMEs and their development

No.	Descriptors	ToR	Final report of evaluation
1.	Ontology (the perspective about reality)	A complex, yet cognizable reality	Extremely complex and dynamic reality. It requires a realistic evaluation and learning approach of type 2 or 3.
2.	Epistemology (the nature of knowledge)	It is intended to deconstruct the structures and reasoning behind the programs in order to improve their knowledge - realism	The evaluation requires several perspectives from several interest groups. Knowledge is subjective and the evaluation process undergoes small changes following consultation with stakeholders - realism.
3.	Methodology	Mixed quantitative and qualitative methods tailored to the context and needs.	Mix of quantitative and qualitative methods according to context and needs

		- Realist approach	-the conditions in the TOR are respected, plus a number of other innovative methods and approaches proposed by the evaluators: for example, Bayesian Network Marketing Survey - focus on theory-based impact evaluation; - Realist approach
4.	Cause and effect	Identification and analysis of causal relationships and change theories underlying the programs are encouraged - realist approach	In-depth causal relationships between context, mechanisms and effects are explored - realist approach
5.	Attitude towards values	There is no explicit reference to values, but the evaluation involves a large number of interest groups	Evaluators do not explicitly refer to values, but they attempt to address evaluation from multiple perspectives and involve a large number of interest groups at central and local level.
6.	Type of approach: top-down, bottom-up	Top-down approach (macro analysis of data) combined with a bottom-up approach	Macro-level analysis combined with micro-analysis, the results of which are included in the

		(carrying out case studies with the involvement of representatives from beneficiary companies, focusing on best practices).	formulated conclusions and recommendations - Realist approach
7.	Stakeholder's involvement	Interest groups are approached in a high number and at many levels, but central ones are more involved - a realist approach	A large number interest groups from both macro and micro levels are involved and their outlook is reflected in the conclusions and recommendations formulated.
8.	Evaluator's role	The evaluator is required to develop classification systems and typologies. He/she has the freedom to make choices in this respect, but does not need to be actively involved - Realist approach	The evaluator's role is similar to those described in the realist paradigm. He/ she deconstructs the mechanisms behind the logic of interventions and formulates assumptions, but does not make known his/her own perspective and value constructions.
9.	The purpose of evaluation and attitude towards results, objectives, implementation process, the role of assessment in	The purpose of evaluation is to ensure financial responsibility, planning / efficiency, to justify and legitimate the resources invested in relation to the results	The aims of improving program planning and efficiency, financial responsibility, improving implementation and acquiring knowledge / learning are to a large

	learning	achieved. However, it is also a matter of acquiring knowledge / learning and improving performance in implementation, through an in-depth analysis, theory-based assessment, and so on. However, the goal regarding the improvement of implementation has a less felt presence than the other three.	extent reflected. And the goal of strengthening administrative capacity is reflected to a lesser extent.
--	----------	--	--

REFERENCES

- Armstrong, Harvey; Wells, Peter. 2006. Structural funds and the evaluation of community economic development initiatives in the UK: A critical perspective, *Regional Studies*, Routledge, vol. 40, nr. 2, pp. 259-272.
- Bache, Ian; George, Stephen; and Bulmer, Simon, *Politics in the European Union*. 2011. Third Edition, Oxford University Press.
- Bachtler, John; Wren, Colin. 2006. Evaluation of European Union Cohesion policy: Research questions and policy challenges, *Regional Studies*, Routledge, vol. 40, nr. 2, pp. 143-153.
- Batterbury, Sarah C.E. 2006. Principles and purposes of European Union Cohesion policy evaluation, *Regional Studies*, Routledge, vol. 40, nr. 2, pp. 179-188.
- Cohesion Policy 2007 -13. 2007. Commentaries and official texts, January 2007

http://ec.europa.eu/regional_policy/sources/docoffic/official/regulation/pdf/2007/publications/guide2007_en.pdf.

- Cohesion policy: Strategic report 2013 on programme implementation 2007-2013

http://ec.europa.eu/regional_policy/how/policy/strategic_report_en.cfm.

- Eser, Thiemo W.; Nussmueller, Eva. 2006. Mid-term evaluations of Community Initiatives under European Union Structural Funds: A process between accounting and common learning, *Regional Studies*, Routledge, vol. 40, nr. 2, pp. 249-258.

- European Commission, DG REGIO, Ex post evaluation of Cohesion Policy programmes 2007-2013, focusing on the European Regional Development Fund (ERDF) and Cohesion Fund (CF) – Work Package Zero: Data collection and quality assessment, Call for tenders 2013CE16BAT060, 2013.

- European Commission, DG REGIO.2011. Monitoring and evaluation of European Cohesion Policy - European Regional Fund and Cohesion Fund - concepts and recommendations. Guidance document, November.

- European Policy Evaluation Consortium (EPEC).2005. Study on the Use of Evaluation Results in the Commission. Final report Dossier n°2: case study reports, European Commission.

- Gaffey, Veronica. 2009. Case Studies in the Framework of Ex Post Evaluation, 2000-2006: Expectations and Experiences of the DG for Regional Policy, European Commission.

- Guba, Egon G.; Lincoln, Yvonna S. 1994. Competing paradigms in qualitative research, în Denzin, N.K.; Lincoln, Y.S. (ed.), *Handbook of qualitative research*, Thousand Oaks, Sage, pp. 105-117.

- Guba, Egon G.; Lincoln, Yvonna S.1989. *Fourth Generation Evaluation*, Sage.

- Guidance document on monitoring and evaluation – European Regional Development Fund and Cohesion Fund – Concepts and recommendations, March 2014

http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf

- Harry N. Boone, J., and Boone, D. A., 2012. Analyzing Likert Data. *Journal of Extension*.
- Hoerner, Julian; Stephenson, Paul. 2012. Theoretical perspectives on approaches to policy evaluation in the EU: The case of Cohesion Policy, *Public Administration*, Blackwell Publishing, vol. 90, nr. 3, pp. 699-715
- House, Ernest R.; Howe, Kenneth R., 1999. *Values in evaluation and social research*, Sage.
- Hummelbrunner, Richard. 2015. Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence, *IDS Bulletin*, January, vol. 46.
- Patton, Michael Quinn. 1996. A World Larger than Formative and Summative, *Evaluation Practice*, vo. 17, no. 2, pp. 131-144.
- Pawson, Ray; Tilley, Nick. 2004. *Realist Evaluation* (document financed by Great Britain's Cabinet)
- Pawson, Ray; Tilley, Nick. 2001. Realistic Evaluation Bloodlines, *American Journal of Evaluation*, vol. 22, nr. 3, pp. 317-324.
- Pierson, Paul. 2000. Increasing returns, path dependence, and the study of politics, *The American Political Science Review*; June; 94, 2; ProQuest Central, pp. 251-267.
- Skolits, Gary J.; Morrow, Jennifer Ann; Burr, Erin Mehalic. 2009. Reconceptualizing Evaluator Roles, *American Journal of Evaluation*, Sage Journals, September vol. 30, nr. 3, pp. 275-295
- t33 srl, ÖIR GmbH. 2015. *Spatial Foresight GmbH, DG REGIO, Ex post evaluation of Cohesion Policy programmes 2007-2013, focusing on the European Regional Development Fund (ERDF) and Cohesion Fund (CF) – Work Package Zero: Data collection and quality assessment, Final report.*
- Westhorp, Guy. 2014. *Realist Impact Evaluation. An Introduction*, Methods Lab, September.

