Minutes of the expert groups

Minutes

Evaluation Network meeting with MS
DG REGIO
01/07/2021, WebEx meeting

1. Approval of the agenda and of the minutes of previous meeting

The Head of DG REGIO B2 Unit Evaluation and European Semester presented the agenda of the one-day webinar. The following agenda items were discussed: the final version of the Staff Working Document for the programming period 2021-2027, big data in the future of evaluation, the relevance of Theory of change in evaluation, and the Italian unit's experience using evaluation evidence using monitoring data. The Members of the Evaluation Network did not add agenda items before the meeting.

Evaluation Network members were informed that the minutes of the meeting of 19 and 20 April 2021 were uploaded on CIRCABC for feedback. No comments were received on the minutes.

2. Nature of the meeting

The meeting was organised for members of the Evaluation Network. Due to its online nature, the participation to the meeting was extended, lifting the limit on representatives from each member state that could join.

The meeting chair encouraged participants to intervene both in the chat and by taking the floor throughout the meeting.

3. SWD on performance, monitoring and evaluation – final draft of the 2021-2027 indicator fiches

The REGIO B2 unit presented the final version of the Staff Working Document on Performance, Monitoring and Evaluation and Indicators descriptions for the programming period 2021-2027.

The objective of the SWD is to provide clarity on provisions and how they are interlinked with one another on evaluation, programming, monitoring using common and specific indicators, and the financial categorisation system.

The adoption process was explained: by the end of the first week of July the SWD and Annexes will officially be transmitted to EU institutions after the adoption of the CPR/Funds regulations on 30/06. It will be disseminated through the Inforegio website, the Cohesion Open Data platform and CIRCABC.

No changes were made to the basic structure of the SWD since December 2020. There have been clarifications of the drafting, the presentation and the referencing including a synthesis of the key changes linked to performance and monitoring, clarification on the meaning of asterisked indicators and on the time of collection of values for results. Additionally references were added to the wider Interreg family of cooperation programmes (IPA and NDICI).

The Unit B2 Team leader for evaluation elaborated on the evaluation and evaluation planning aspect of the SWD. The text was developed based on experience from the 2014-2020 period thanks to minor legislative changes that consisted mostly of simplification. The document is practical and leaves methodological questions for further discussions, for which the Evaluation Network is a good forum. The first challenge in this period is to submit the Evaluation plan. The Helpdesk is available for active support in elaborating the plans.

The Unit B2 Team leader for monitoring presented the indicators and monitoring aspect of the SWD. The pre-drafts of indicator description in the form of fiches, were provided in CIRCABC on May 31, including the anticipation of final legislation and new indicators in "clean" and "track changes" versions. Work has continued to finalise and eliminate inconsistencies post the Commission's inter service consultation.

Some legislative changes impacted on the indicators. There were new specific objectives and the transfer of specific objectives, more indicators were given asterisks that allow use across SOs and POs. Improvements were made to the data transmission template for the reporting of results to make it more intuitive.

Within an overall objective to minimise changes to the fiche already circulated the changes made to the indicators include the following: new fiches for 7 indicators from trilogues discussions, some integration of comments from the member states, changes to the measurement units and definitions, harmonisation of terms used across indicator descriptions, count among the changes taken on board.

Some comments received were not taken on board: where the comments were not in accordance with the indicator names listed in draft legislation or agreed in the initial Council compromise; comments linked to specific national circumstances or specific project circumstances; comments addressed at the seminar on transport result with JASPERS and the seminar on estimating GHG emissions with EIB.

Regarding categorisation, major revision were made to intervention fields to align the RRP and cohesion policy systems. The changes touched particularly clarifications on the use of higher climate weightings. A new dimensions on gender equality was added in Trilogues.

Apart from the SWD, B2 is in parallel building with IT colleagues the SFC 2021 programme template, developed with drop down lists to support encoding of specific objectives, common indicators and intervention fields.

Reference metadata tables will be launched early July on the Open Data Platform, and the first reporting of implementation early 2022. There will be a higher rhythm of reporting in the future.

On Inforegio the SWD and Annexes in PDF will be published, as well on CIRCABC along with excel tables on indicators and categorisation. On the Open Data Platform the metadata datasets on common ERDF/CF/JTF indicators and cohesion categorisation will be published as well as data stories.

NB: The materials are since published on this Inforegio webpage: https://ec.europa.eu/regional_policy/en/policy/evaluations/guidance/#1

4. Big Data in the future of evaluation

Mr. Frans Leeuw (Evaluation Helpdesk) began by presenting the basic concepts of Big data, Artifical intelligence and machine learning, and big data ecosystem. Big data needs analytics. Statistical programs and software including machine learning and artificial intelligence are often applied. AI applies different types of machine learning like supervised and unsupervised learning and deep learning. Supervised leaning involves machine learning algorithms that learn under the presence of a supervisor or guidance. Unsupervised learning uses machine learning algorithms to analyse and cluster unlabelled data sets, which algo's discover hidden patterns in data without the need for human intervention. Deep learning has been specifically modelled after the human brain. It works with algo's of a brain like logical structure, called artificial neural networks. This all represents the big data ecosystem.

Mr. Leeuw's approach to big data and evaluation is two sides of the same coin illustrated by the two following questions: what have BD/AI to offer to evaluations of policies and programs? And how to evaluate the use of BD/AI in policy programs and interventions?. To explore the first side, Mr. Leeuw presented big data tools such as social media or radio call-in programs and relevant examples of evaluation applications such as respectively opinions and sentiments and identifying sources of conflict. He then developed more on topics related to the EU and Member States programs and activities. With some examples on mobility and infrastructure, auditing and inspecting neighborhoods, digitalised educational assessments, social enterprise networks, and textmining, natural language processing.

There are some bias regarding pitfalls, risks and challenges using Big Data/AI: the first one being that it only includes subjects using the apps known as the digital divide issue, the second bias being legacy problems contributing to social bias (racial, economic, gender), the third being that gathering the data comes before the problem.

The second side of the coin refers to the challenges and rapid diffusion of AI/BD in society. Mr. Leeuw elaborated on a three-step approach to evaluate the use of BD/AI in programs, demonstrated through an example on medicine. The first step being to specify the problems that have to be addressed, the second to search and articulate the assumptions underlying working with BD/AI: the theory-driven approach, the third to distinguish between evaluating AI algorithms as such working within and between computers and evaluating AI algo's implementation in practical situations like health care, education, transport, justice etc.

In conclusion, Mr. Leeuw noted evaluators pay increasing attention to AI/BD. The focus is more on description/ stock taking than on contribution or causal impact analysis. Working with BD/AI is confronted with also other pitfalls, risks and difficulties. These make that evaluating how BD/AI work out in practice and in combination with Human Intelligence is very relevant indeed.

Discussion:

Unit B2: I think the BD characteristics presented in the beginning define their usefulness for evaluation. I would think there is an issue there as to who is qualified to judge these characteristics. Do such profiles exist in the evaluation community? My second question is: are there any low hanging fruits that are easy to use in practice?

Mr. Frans Leeuw: It is indeed relatively complicated. What we recommend is to bring people together with different expertise. Evaluation is itself an interdiscipline. As for the low hanging fruits, there are new types of content analysis such as textmining.

Mr. Terry Ward (Evaluation Helpdesk): Are we not in danger of creating more blackboxes with BD which relies very much on AI to interpret relationships which exist within the data? How to overcome this specific problem.

Mr. Leeuw: I have found in the literature one approach to the opaqness, which comes very close to what we as evaluators are doing in theory-informed evaluation.

Unit B2: One take away for me is the need to try and bridge the gap between the academic literature to the tasks our colleagues in the Member States have, which is to try and evaluate investments. Which to a large extent has been rather difficult. We will expand on this in our upcoming conference in Portugal.

5. The importance of Theory of change in evaluation

Drawing from his experience at the Evaluation Helpdesk, Elliot Stern presented his lesson learned through Helpdesk support with both Member States and Commission-led ex post evaluations in the form of Work Packages.

The speaker discussed Theory of change as the foundation of all theory-informed approaches. Originally traditional evaluations avoided theory as they were to identify dimensions of success and measure success or failure. Since there has been a theory explosion in evaluation with emergence of Theory of Change, Programme Theory, Theory-based evaluations etc. Which have been used to explain causal connections between policy interventions and intended results. The main three drivers to this explosion appear to be that policy priorities and how priorities are delivered have changed, there is increasing interest in results in making a difference, finally that we are less tolerant to policy failure.

There are some key challenges to the establishment of theory use in evaluation. The first of which is language confusion that is based on history. Mr Stern used a river allegory to explain evaluation, its different streams being intervention logic, programme theory, log frames, impact evaluation etc. The second is how we are still stuck in brand loyalty to common principles. The third is practice convergence, where all theory informed approaches rely on common principles: causal

mechanisms, causal pathways, and contextualisation. The fourth challenge is a capacity one, in terms of skilled contractors able to deliver what they promise, data sources that are beneficiary and intervention relevant, flexible management capacities and procedures and coping with misaligned timescales.

To conclude, Mr. Stern stated the following take-away messages. Theory informed approaches are developing slowly but there are some evidence of convergence in language and practice. Theories of change are the foundation of any kind of theory-led evaluation, and are most times enough. Getting this right at the managing authority level would help a great deal when completing the ex-pot work as it would allow for a library of theories of change instead of totally reconstructing on the back of very weak foundations. Answering how and why questions make sense when these answer stakeholder questions, granted the importance of maintaining dialogue. Benefiting from theory-informed approaches requires decent data and understandings of what is needed to make data sources more case-relevant and aligned. Finally keeping in mind that while completing evaluations, one will continue to face capacity problems.

Discussion:

Unit B2: We as commissioners are facing similar challenges to Member States. It is equally important to answer 'how' and 'why' in particular for policy-learning purposes and communication purposes, as it is to answer 'what'. We are starting to consider how to look into the impact of COVID, and how it has changed the initially defined objectives of the programmes. This we will dedicate another evaluation upon completion of the ex-post.

SE: There are so many things to do in preparing for a programming period. Typically we have processes to develop the application forms for the projects. The reporting templates the projects will use. We have more time to finalise the evaluation plans whereas the data needs to be decided on at an early stage. When the evaluation starts we realise there are data that should have been included from the start.

Unit B2: We are all relying on somebody in terms of getting the data. It is very important to recognise the role of each and everyone in the process and try to be as helpful as possible to the whole chain. Data is the base for everything in terms of policy learning.

Mr. Terry Ward (Evaluation Helpdesk): Thinking about the importance of data: when it comes to identifying to data that needs to be collected for an evaluation, my view has always been that evaluation or evaluation planning should be an integral part of the programme design process. At that stage, an embryonic theory of change should be spelt out and the data required to understand a subsequent of that theory of change should be identified. This is an ideal situation that is rare to find in practice. Data issues emerge at the same time as an evaluation is being planned. The importance of a theory of change is irrespective of what kind of evaluation will be undertaken. A good counterfactual evaluation has to have at its basis a theory of change.

Mr. Stern: People have got different tools in their toolbox. If what they do is counterfactual they will always do counterfactual. I agree that every evaluation should look at whatever approaches

are relevant to answer evaluation questions that are being asked. The leading point to an evaluation is what do we want to know.

Mr. Ward: I think there will always be a need for surveys to understand what initiates behaviours. In time it is important to have the quantitative data underlying that so that you can at least in some sense test whether the qualitative information you are getting is valid in terms of what the quantitative data is showing. A good impact evaluation needs both quantitative and qualitative data. It is the quantitative that is important to plan ahead.

Unit B2: Unless you decide to recourse to artificial intelligence interpretation of big data to supplement for the surveys.

RO: The problem of low quality ToC and not enough and relevant data in the monitoring system might be due to the use in the programming of a very simplified form of intervention logic, which is reduced to input-output-outcomes-impacts.

Mr. Stern: There is a transition from an intervention logic that is often input output to preparing or reconstructing a theory of change. Even if we have an intention objective driven intervention which is more outcome output input type of thinking. If people think that they are having to stay within the boundaries of the starting intervention logic, that is a constraint. This highlights the need for a theory of change that has many more factors in it, which takes account of the context and the pathways.

Mr. Leeuw: I completely agree with the point of doing a deeper state. There is a already a word for used in academic work "theory knitting".

Unit B2: Sometimes what went wrong and the failure of the evaluation is the most important lesson learnt. This is very important for us and the managing authorities.

Unit B2: It is not necessarily only to learn from our own mistakes, but as well common ones so that nobody commits them in different circumstances.

Mr. Stern: One of the functions of the Helpdesk is to act as a mirror or a reflection. It is a continuous process of learning, where both failures and successes have teachings.

6. Use of Evaluation Evidence and Monitoring Data in an Inter-organizational Environment: Italy's Evaluation Reviews and Labs

The Italian member of the Evaluation Network discussed Cohesion Policy as one of the most evaluated, monitored and open policies in Italy, with an emphasis on how to put evidence and data at work. She first gave an overview of the evaluation system in Italy that is decentralised. Evaluation units are present in each regions and in most of the ministries involved in Cohesion Policy. The network meets once a month and works in work groups. They perform two tasks of actions: common activities in which the network supports what each evaluation unit does for its organisation, and the collective activities where participants together produce knowledge for the Cohesion Policy community.

The presentation then focused on how evidence was provided to Managing Authorities for the 2021-2027 programmes. In 2019, evaluation units were involved in reading and learning activities. In 2020-2021, Managing Authorities were involved in actually using knowledge to support the choices that would be done in programming for the following period. Collective intelligence is at the heart of practices, allowing for bringing knowledge close to intended users, involving public sector in processing evidence and data.

The evaluation reviews project produced six non-systematic narrative synthesis, learning from evaluations. A simple methodology was used to find, collect and process evidence from Ministry or regional authorities, SNV observatory and other sources. A complex one was chosen to facilitate the collective work in an inter-constitutional and non-hierarchic environment. Exchange of thought, information and experience was at the center of this project.

Three labs were then organised involving Evaluation Units and Managing Authorities to use knowledge for 2021-2027 programmes. The methodology used was based on the requests from the programme template. For each lab, there were an overarching question from the template. The architecture was slightly more complex as each lab had its central group made up of coordinating units: the evaluation unit and the sector unit responsible for interventions, for rural development, and research institutes on the environment. Some regions agreed to participate in more than one lab. Very wide involvement was achieved, and made for wider reach of results and knowledge overall.

7. AOB

Unit B2 provided updates on the Evaluation Helpdesk as follows. Surveys were sent for the summer school and methodological support, so far 12 Member States had responded. It seems like most MS look for support in defining evaluation plans for the upcoming period but there is still room for more support request. Most respondents want to learn about evaluation design and methods. Regarding thematic point of views, respondents prefer to concentrate on TO1, 4 and 9. There is a need for more input to define the programme for the summer school and plan methodological support for the following year. Therefore the deadline has been extended by two weeks to collect more input.

Unit B2 presented the updates regarding the upcoming Evaluation Conference. At this stage, more respondents would join the conference online. At least one seat reserved for each Member States, for which guests are invited to register as soon as possible and encode your preference.

8. Conclusions

On 16-17 September 2021 will take place the 9th Conference on the Evaluation of EU Cohesion Policy, in Porto, Portugal.

Unit B2 thanked Evaluation Network members for the very constructive sessions and debate, asking to continue to sending questions on evaluation that can be discussed during the next meetings.

9. List of participants

The Evaluation Network meeting with Member States that took place on 1 July 2021, was attended by 67 governmental experts from 27 Member States, 1 expert from the Evaluation Helpdesk contractor APPLICA, 2 experts from INTERACT Office Vienna, 2 external speakers, and 14 REGIO Unit B2 colleagues.