8th international evaluation conference

EVALUATION post 2020: evolution or revolution?
THE 8TH INTERNATIONAL INVESTMENT EVALUATION CONFERENCE „EVALUATION POST 2020: EVOLUTION OR REVOLUTION?“, WHICH TOOK PLACE IN VILNIUS, WAS ATTENDED BY OVER 120 PARTICIPANTS FROM 19 COUNTRIES. THE ORGANISER OF THE CONFERENCE IS THE MINISTRY OF FINANCE OF THE REPUBLIC OF LITHUANIA. THE CONFERENCE AIMS AT ENHANCING EVALUATION CAPACITIES. THE OBJECTIVE OF THE EVENT IS TO CREATE AN INTERNATIONAL PLATFORM FOR SHARING BEST PRACTICE AND IDEAS WHICH SHOULD ENCOURAGE NETWORKING AND RECIPROCAL LEARNING.
“More than one hundred evaluations of European Structural and Investment Funds’ (ESIF) interventions have been carried out since 2004. They provided over one thousand recommendations. Most of them have been implemented, which resulted in significant improvements in the design and implementation of the interventions. Perhaps more importantly, evaluations have gradually changed the decision-making culture: evidence is increasingly gaining currency at the expense of inertia, special interests and ideologies.

First, in the words of American scientist Aaron Wildavsky, ‘the owl of wisdom always arrives too late’. Policy makers need the evidence generated by evaluations at specific points in time – when setting objectives and targets, designing and improving interventions and allocating resources. The evidence on ‘what works’ is in high demand before launching interventions, not a couple of years later. Yet most evaluations focus on already ongoing or past interventions. While they generate ample evidence, the knowledge arrives too late. Hence, we should discuss how to streamline evaluations so that timely evidence feeds directly into the decision-making process.

The second challenge concerns transforming evaluation into a system of knowledge production and management, rather than a system of producing reports. Clearly, decision-making and evaluation cycles are not always operating in synchrony. Hence, how can we link the needs of policy makers with the evidence produced by evaluations? How can we ‘store’ the knowledge produced by evaluations and effectively ‘retrieve’ it, once the demand for it peaks? Who should be the ‘knowledge brokers’ within the evaluation and decision-making communities?

The third challenge and opportunity is presented by the emergence of new technologies, in particular, artificial intelligence, predictive analytics and similar innovations. Computer algorithms ‘know’ with great certainty what books you are likely to buy, where you are likely to travel and which social circles you might want to connect with. Can we use these technologies for the advancement of public good? How could these technologies help us to design and implement better interventions? If artificial intelligence is likely to solve our major problems, what is then the role of expert knowledge offered by the evaluation community?”

I would like to use this opportunity to outline three challenges that are particularly important for the policy makers across the globe.
With this presentation, the Commission outlined its plans on the improvement of evaluation of Cohesion policy during the programming period 2021-2027. The changes aim at balancing the need to consolidate the achievements made in 2014-2020 on the one hand and taking into consideration stakeholders' proposals on the simplification of the regulatory framework on the other hand. “Our policy is targeted at efficiency and results; in order to achieve good results, we had to develop complex and sustainable programmes. Negotiations with EU experts are currently underway, we want to specify the list of indicators, because it is extremely important to explain clearly how to create adequate policy. We concentrate not only on effectiveness and efficiency but also on compatibility. We also expand our ways of effective handling of evaluations; for instance, they will have to be publicised on the internet,” - Ms. Hristcheva said at the conference.

She states that the amount of funds for evaluations originally provided for in the EC proposal has decreased by 20 % down to 273 billion Euro, however this remains a significant amount.
During the last three decades, evaluation has established itself as an important tool for evidence-based policy making all over Europe. At the turn of the century, the EU, OECD and the UN have helped spread evaluation over the whole world. In Europe, the European Structural Funds gave an important impetus for the spread of evaluation and for the fast development of evaluation culture, especially in Central and Eastern European countries.

At the conference, Dr. Meyer presented the research project launched in 2016 by the Saarland University Centre for Evaluation CEval on the institutionalization of evaluation in Europe; the results of the research are about to be published in a book called “The Evaluation Globe”. By using a comparative research design, more than 25 authors from 18 countries present the best examples on implementing evaluation in the political system, the society, the academia and research institutions, offering the broadest overview on the institutionalization of evaluation in Europe ever published.

This presentation reported on the comparative results of the research and addressed the question whether the institutionalization is already an indicator of professionalism and the rise of the European evaluation culture.

“We observe the bottom-up trend of the evaluation process and institutionalisation, i.e. the processes start at the level of the stakeholders of the sector and develop rising up to the legislative level. In general, evaluation results are not so widely used as the results of public surveys, while the civic society gets involved only when they see the actual benefit,” – Dr. Meyer said.
The presentation explored the probable shifts of the evaluation system determined by three pillars. First, the traditional objectives of conducting evaluations (i.e. ensuring accountability and improving public policy) are based on the assumption of supremacy of knowledge and evidence over ideological attitudes in the formulation and implementation of public policy. Yet the aftermath of the global financial crisis witnessed the growing polarisation within societies, which has raised doubts as to how long this assumption is going to remain acceptable. Public policy is increasingly becoming a tool for ideological battles: how will this change the nature of impartial/technocratic evaluations?

Second, evaluations are driven by the theories of change that aim at establishing causality among interventions and behavioural changes of individuals/organisations. These theories are used both in order to explain the failures of the past and to make informed guesses about the future results of the programmes. Nevertheless, the increasingly prevalent models based on big data do not refer to the causality analysis in order to guess the future: how can this change the nature of evaluations?

Third, traditional evaluations rely on a limited number of data collection methods, such as interviews, surveys, monitoring data, etc. “All-encompassing digitalisation provides access to a huge amount of data at a low cost. What does that mean? It means that we can do real-time evaluations,” – Dr. Martinaitis said at the conference.

Trend no. 3: Big data

Evaluation relies on inherently limited and expensive-to-obtain data:
- Interviews, surveys, monitoring data, and similar

Big data:
- Exponential growth – approx. 90% of data has been created during the past decade
- Volume and variety: n= (almost) all, individual level, time series
- Cheap: large initial costs of setting-up the system, but insignificant marginal costs of additional data collection
Dr. Gustav Jakob Petersson discussed methodological challenges and trends of future evaluations. In his presentation, he raised the question what the best methods to identify the true impacts of an intervention are. These questions seem to have been the most heavily debated ones among evaluators.

The presentation discusses the opportunities and challenges within the evaluation context, such as the introduction of big data analytics, systematic reviews and syntheses, as well as complexity theories.

The analyst makes the following statement: when discussing the benefits of such methodological innovations, it is important to bear multiple causality theories in mind. This may help avoid the bias in the evaluation of methodological innovations and overly polarised discussions.

“We should concentrate on two issues. First, the research design [i.e. what an impact evaluation should look like], and what it means to us, i.e. what do we want to find out when ordering an evaluation. Second, we should not be talking different language as we are all in the same boat; we live in the age of big data; i.e. only one quarter of global data was kept in the digital format in 2000, while nowadays, 99.9 per cent of the data has been digitalised. Therefore, we need to encourage cooperation among institutions,” – Dr. Petersson said.

Integration through synthesis?

- Meta-analysis (in a particular form...) vs. (for instance) realist synthesis
- Ex. combining meta-analysis and implementation studies: “Meta-modeling Social Programs: Methodological Reflections on a Practical Application” (Lemire 2017)
“Why should policy evaluation capacities be enhanced? Because higher capacities mean a higher quality policy cycle, they prevent unilateral policy making, they enable people better distinguish the truth within excessive amounts of data,” – Stephane Jacobzone said at the conference.

The presentation outlined the main trends across OECD countries for framing evaluation practices and systems. It discussed the main institutions in charge of evaluations and the challenges they encounter when performing them and using their results, as well as the extent to which evaluation practices are defined in legal and regulatory frameworks. It also discussed the types of quality assurance mechanisms used by OECD countries and the tools that can be used to promote evaluation and the use of their results, including skills upgrading for civil servants. “We are now conducting the research in 42 countries, whereby we seek to clarify the purposes for which the countries use evaluations and whether they refer to their results in their actions. The research has shown that the countries seek to make the evaluation system inclusive, however we encounter the challenge of the missing strategy of policy evaluation and the lack of political will to evaluate policies,” – Mr. Jacobzone said at the conference.

He believes that there are many quality assurance mechanisms, however almost 40 countries do not apply any mechanisms for the measurement of contents quality: “How do we promote better evaluation? First, by training, i.e. by evaluation networks, experience sharing.”

The participants discussed the impact of the tighter rules of personal data protection on the situation, the challenges they create for the evaluation process, what is needed to receive the data necessary for evaluation, who and for what reason can have access to data, and how the data should be provided.

Integration through synthesis?
The speaker discussed the evaluation of the training funded by the ESF and analysed whether new data collection strategies proved effective. The European Social Fund (ESF) has provided more than 1 billion Euro to support human capital development in Lithuania for the funding period 2014–2020. The major part of these ESF funds goes to the funding of various training activities. However, notwithstanding the scope of the resources allocated, the evaluations of training interventions are usually based on only subjective quality data. Our evaluation seeks to challenge this traditional attitude towards training evaluation.

“In order to estimate the benefits of training, we have opted for an innovative method rather than the traditional one: we have interviewed all beneficiaries. We have interviewed to nearly 200 000 beneficiaries; approximately 100 000 of these had e-mail addresses,” – Mr. Gaušas explained. The speaker discussed the merits and challenges of three data collection innovations that were applied during the evaluation: 1) the analytics of free jobs on the market revealing how real-time labour market monitoring system can be applied for the assessment of the relevance of enhanced competences as a result of labour market training activities; 2) the advantages and disadvantages of real-time monitoring of training; 3) the results and shortcomings of the large-scale survey sent to all beneficiaries of training activities.

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>More “objective” data [compared to interviews]</td>
<td>Limitations determined by the chosen logic of the categorisation (e.g. whether driving courses should be categorised as professional training)</td>
</tr>
<tr>
<td>Continuous process of data collection: identification of trends</td>
<td>Specific tools and constant maintenance are needed to scrap the data</td>
</tr>
<tr>
<td>Flexibility, depending on the question (e.g. labour demand by the level of education in LT regions)</td>
<td></td>
</tr>
</tbody>
</table>

Use of big labour market data
In her presentation, Dr. Astrid Molenveld analysed how evaluations can change the results of public intervention. The changes in the area of evaluations may be described by a few trends. Discussions on how to conduct policy evaluation are increasingly about numbers, algorithms, big data, evidence, and evidence-based interventions. The view on policy evaluation is ever more “quantitative”, both on ex-ante evaluation (e.g. the evidence bases to substantiate policy interventions), as well as on ex-post evaluation. When we think about evaluations, we simply cannot make an estimate of all the ways to make an intervention effective or grasp its effects to the full extent. An accumulation of all available evidence about interventions is a good start of a possible intervention, however there are other factors at play. “We have noticed that the public administration model has its shortcomings, i.e. the results do not reach the parliaments or governments. It is difficult to find the ways to shift towards a cooperative inclusive culture. Our aim is to generate the results suitable to many multifaceted institutions, to create common value, and this should happen externally among organisations rather than internally within an organisation,” - Dr. Molenveld shared her experience. She believes that the context - especially in the public sector - matters a lot: it is important to have public support, mobilise intervention advocates, etc. She gave an example about the evaluation of the housing renovation programme in South Rotterdam, a poorer city district (The National Program Rotterdam South).

Dr. Molenveld believes that the task for researchers and practitioners who are interested in further development of the evaluation field is to look for alternative methods for policy evaluation, which smartly combine data, evidence and leave room for reflexivity and learning. New types of evaluation which can assess interventions and public outcome are being developed (e.g. SROI and QCA) however there is still much work to be done.

### Qualitative Comparative Analysis (FsQCA)

<table>
<thead>
<tr>
<th>Broad support</th>
<th>Facilitative leadership</th>
<th>Institutionalization</th>
<th>Shared finances</th>
</tr>
</thead>
<tbody>
<tr>
<td>- links with many institutional partners</td>
<td>- Convenors</td>
<td>- contracts/convenants</td>
<td>- Shared funds, co-financing</td>
</tr>
<tr>
<td></td>
<td>- Catalysts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mediators</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Sorensen and Torfing (2012:8)</td>
<td></td>
</tr>
</tbody>
</table>
Composite programmes pose new challenges for M&E (monitoring and evaluations); the insights and methods from the systems field might be useful here. Based on an overview of these challenges, Mr. Hummelbrunner discussed three issues in his presentation: 1) he outlined the rationale for using systems thinking, explained the three core concepts (“interrelationships,” “perspectives” and “boundaries”) and gave some guidance on choosing appropriate systemic methods; 2) he analysed the use of systems thinking when monitoring composite situations, where monitoring, being based on predefined quantitative indicators, is inappropriate or insufficient; 3) he discussed the implications of systems thinking for (impact) evaluations.

In his presentation, he was referring to professional practice, notably the book “Systems Concepts in Actions. A practitioner’s toolkit”. “What is monitoring, which is so important for us? We observe recurring indicators; however, they are often insufficient for us to make an all-encompassing evaluation of a subject. We need a more flexible, more dynamic system in order to analyse the impact, causality and evaluate external factors. When you have the data, don’t try to look for justification via them, look for exceptions. Such monitoring is learning,” - Mr. Hummelbrunner said when sharing his experience.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Characteristics</th>
<th>Causality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>• high certainty and agreement&lt;br&gt;• known right answer&lt;br&gt;• best practice ‘recipes’</td>
<td>• clear, predictable and controllable</td>
</tr>
<tr>
<td>Compli-</td>
<td>• some uncertainty and some disagreement&lt;br&gt;• good practices&lt;br&gt;• requires analysis, coordination and expert knowledge</td>
<td>neither obvious nor predictable&lt;br&gt;depends on context&lt;br&gt;alternative routes</td>
</tr>
<tr>
<td>cated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>• high uncertainty and high disagreement&lt;br&gt;• every situation is unique&lt;br&gt;• requires observing relations and (behaviour) patterns</td>
<td>only evident in retrospect&lt;br&gt;depends on initial conditions</td>
</tr>
<tr>
<td>unknow-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>able</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Glouberman and Zimmerman, Kurtz and Snowden, Rogers, Patton
In her presentation, Ms. Šemetienė analysed the inexhaustible and periodically recurring subject (or perhaps a problem) which has been of interest to the evaluation community for many years: what are the actual results of an evaluation, how can the use of evaluation results be measured, how can the impact of evaluations be determined? She discussed the understanding of evaluation results and of their use: she framed it as a recommendation and its implementation and/or generated knowledge and its adaptation to the changing fundamental beliefs of the interested parties, the broad consensus on the fundamental changes and the empowerment of the owners of the change.

The speaker raised the question whether the narrowed use of evaluation results focusing on only the implementation of recommendations really reveals the actual use of evaluation results, and whether we are not sacrificing the content to the form.

“Strategic proposals are most often of high quality and they are normally accepted by the institution receiving them. However, over a quarter of proposals have no “owner”, i.e. ministries cannot instruct governments on proper decision-making. The knowledge generated during evaluation gets frequently lost as experts rather focus on analysis, leaving insufficient time for a consensus on the implementation of strategic proposals,” - Ms. Šemetienė said when discussing the arising challenges.
The planning of public policy, as well as the strategic decisions of cohesion policy, are exclusively aimed at the economic and social convergence. In strategic planning documents, implementing programmes and evaluation reports, this aim is normally reflected in macroeconomic indicators, such as the GDP, employment rate, labour productivity, income, assistance. The measurements of life quality aim at expanding the concept of social and economic development, and the condition of social development is measured through the attitude of individuals towards their life quality. The presentation discussed the concept of life quality measurements and the totality of new knowledge/evidence generated by the life quality measurements used for the planning and evaluation of public policy. “The evidence collected during evaluation help decide whether an intervention, targeted at specific results, proved effective. Therefore, in order to be aware of the relationship we will be looking for, we need to define the deliverables of an evaluation,” – Mr. Jatkauskas said at the conference.
Ms. Bruno delivered the presentation “Learning from Evaluations to Shape the Future: Developing Key Impact Pathways for the European Framework Programme for Research and Innovation.” In June 2018, the Commission submitted a proposal for the establishment of Horizon Europe, the ninth European Framework Programme (FP) for research and innovation with a proposed budget of nearly EUR 100 billion for 2021-2027. Since 1984, the EU investments in research and innovation FPs contributed to a large number of scientific advancements and discoveries which generated significant benefits for the society and the economy. These impacts have been documented in multiple evaluation exercises and dedicated studies; however, such evaluations still face common methodological challenges and limitations. The biggest difficulty is to identify and capture the direct and indirect long-term effects that can be attributed to these risky investments in composite and open research and innovation systems.

“It is difficult to prove that research and innovation are extremely beneficial. We face the challenge of limited data supply and problematic collection, and of an adequate evaluation timespan. Scientific, social, economic and technological effects are long-term, while we cannot afford to wait for 25 years, therefore we have to clearly identify the problem and make a decision on evaluation criteria,” – Ms. Bruno said.

Based on the lessons from previous evaluations, in order to monitor progress almost in real time, the European Commission proposed a revamped monitoring and evaluation framework for Horizon Europe built around a set of impact pathways.
Dr. Bastiaan de Laat’s subject at the conference was “Understanding Policy Impact”. OECD Committees give strategic guidance for work programmes and are responsible for ensuring that most of the work of the organisation is performed and the results are made known to the public. There are currently around 30 “Part I” Committees, and, taking into account their substructures, the OECD counts well over 200 of its component bodies in a great variety of fields. Taken together, the Committees issue around 250 documents per year and organise numerous meetings, fora, workshops and other events. The speaker stated that some OECD documents have a binding character (e.g. international agreements such as the Anti-Bribery Convention or OECD decisions), however most of the time, for example in the case of recommendations, declarations or indicators, they lack a formal obligation for member countries to use them in national policy-making processes. Yet the expectation exists that the OECD work contributes to policy making, and it may do so in different ways: by providing supporting data or evidence, offering new principles, guidelines or models for policy making, by giving possible directions for policy reforms, etcetera. Moreover, the OECD work performs an important benchmarking function between countries and facilitates policy reviews. It is expected that policy makers in member countries are aware of the relevant OECD products, use them for their work, and that this ultimately leads to some sort of policy impact or contributes to policy making in other ways. The OECD In-depth Evaluation Unit seeks to assess the relevance, effectiveness and efficiency of the work done by the OECD Committees, which is to a great extent difficult to measure objectively.
The speaker presented the results of the evaluation on the financing of Lithuanian economic sectors after 2020; the evaluation has been commissioned by the Ministry of Finance of the Republic of Lithuania. The scope of the evaluation covers public interventions in 14 public policy areas: education and science, energy efficiency and housing renovation, business, culture and tourism, public administration, information society, environment, employment and social inclusion, health, regional development, research and innovation, transport, energy, demography and migration.

The overall objective of the evaluation is to optimise public interventions in the said policy areas, in order to ensure sustainable growth of the society’s quality of life in the medium- and long-term perspective. Three tasks have been set in order to achieve the objective:

1. To identify the structure and the scope of public interventions in the above-mentioned public policy areas by type of intervention (regulation, financing, production, income support) and funding sources (state and municipal budgets, EU or other support funds) and to evaluate the relevance, sufficiency, coherence and effectiveness of all types of interventions in achieving the public policy objectives;

2. To examine the compatibility and relevance of public areas (objectives) in seeking to achieve sustainable growth of life quality for the society in the medium- and long-term perspective;

3. To identify the potential directions for decreasing the dependency on EU Structural Funds (EU SF) in 2021-2027 and to assess the possibilities of replacing the EU SF investments in public policy areas with non-financial interventions by the state or private funds in order sustain economic development and social welfare.

The evaluation assumed that one of the main problems of EU Structural Funds investments (and of all investments in general) is insufficient coordination of investments with other types of interventions, including regulation, tax policy, communication, etc. The assumption relied on the previous evaluations carried out in Lithuania in the areas of poverty reduction, waste and competitiveness. The preliminary results of the evaluation highlight the importance of both the coordination of the different types of sectoral interventions and of inter-sectoral coordination of interventions. For example, improvements in the quality of public services such as education, health and culture require faster optimisation of the network of service providers, which can lead to cross-sectoral learning. One of the main preliminary conclusions is that the main obstacles to Lithuania’s economic growth, such as poverty, inadequate skills of the labour force, insufficient level of innovation and increasing regional differences, are intersectoral, while Lithuanian public policy lacks effective interinstitutional cooperation instruments to address these challenges. “Why is coordination absent? This is because the funds are separate systems and policy making is not integrated. Coordination of interventions is a success factor; coordination might be slow, however when interventions are coordinated during policy making, communication among institutions improves,” – Mr. Maniokas said.
The participants of the discussion debated whether more efficient use of interventions is possible. The Vice-Minister of Finance of the Republic of Lithuania Miglė Tuskienė stated that the ways to increase efficiency do exist, because evaluations change the decision-making process: “I will give you an example. We are carrying out a budget reform and we intend to implement it in 2021-2023. The reform is based on the evaluation of Structural Funds interventions.”

The Vice-Minister of Economy of the Republic of Lithuania Marius Skuodis raised the question what resource is most scarce in policy making. “I believe that the most missing element is the focus. We are urged to do everything very quickly, while evaluations take a while, and we need to spend some time on analysis and assessment,” - he said.

Dr. Bastiaan de Laat urged a discussion on whether evaluation results should be strictly treated as the only facts to be taken into consideration, or whether evaluations are a way to invite a discussion on more effective public policy and decision-making processes.