

NIRAS International Consulting

NIRÁS



Petra Mikkolainen

Senior Consultant, NIRAS International Consulting

Petra's background is in the implementation and evaluation of development cooperation and policy in multiple sectors. Her current areas of interest include the interconnections between evaluation and foresight and futures literacy in supporting organisations' and companies' strategy work. In past years, she has taken courses on futures literacy and foresight with the Institute for the Future and the Finnish Future Fund Sitra. In December 2020, she acted as a core team member in implementing a Futures Literacy Laboratory on the Future of Evaluation in Society. The event was organised jointly by Unesco, the European Evaluation Society, and NIRAS International Consulting as part of Unesco's High-Level Futures Literacy Summit.

🕅 petra.mikkolainen@niras.fi

in https://www.linkedin.com/ in/petramikkolainen/ Evaluation - by definition - assesses past events to give recommendations for future action. There is an underlying assumption that what has (or has not) worked in the past will also work (or will not) in the future. In other words, it is supposed that the context in which the past events occurred will remain the same. This idea seems problematic in the current world, where volatility, uncertainty, complexity, and ambiguity (VUCA)¹ are the new normal. One solution is to integrate methods of foresight into the evaluation project cycle. This idea of combining evaluation and foresight is relatively new and untested in the sector. This discussion paper proposes ways this integration can be done in practice in different steps of the evaluation project cycle.



NIRÁS

The absence of future in evaluation project cycle

A central question in many evaluations is: what should the institution do in the future to improve the policy/project design and its implementation? Indeed, evaluations have responded to that question – often usefully – but the response is constrained because the evidence informing it is only backwards-looking.

In this blog, we will unpack some of our suggestions on how to mainstream futures thinking into evaluations. We walk you through the main steps, from defining the study's purpose to formulating recommendations.

Some definitions

Evaluation assesses the merit, worth or value of a policy, project, or programme. It uses information from the past to make recommendations for the future. Foresight (also called strategic foresight) explores mid to long-term futures and drivers of change to inform policy and strategies. These are short definitions; both concepts include more nuances and related terms.

Our blog **"Start here! Evaluation and foresight – your quick guide to basic concepts and terms"** provides a concise introduction to both "worlds".



The door icon indicates the **introductory** section that prepares the reader for the topic.



The signal icon highlights **examples** of institutions or assignments where future-oriented thinking has been integrated into evaluation.



The bulb icon leads the reader to explore **ideas** of how the integration of futures thinking and foresight could be implemented in evaluations that operate under real-world time and money constraints.



We also pose some **questions** that we consider relevant. You are more than welcome to contribute to the discussion.

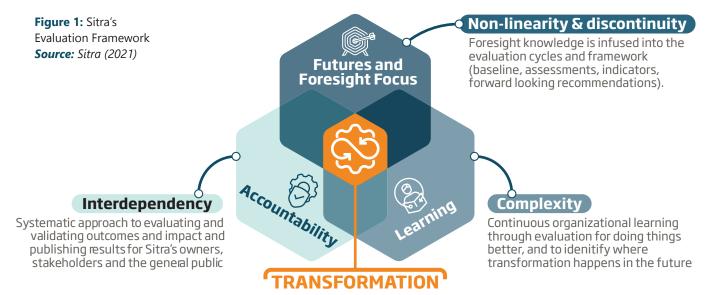


NIRÁS

1. Evaluation purpose and objectives

In a summative evaluation, the evaluation purpose and objectives might relate to understanding the successes and challenges of the intervention (accountability). In more thematic or strategic evaluation, the focus is often more on learning. Then, we often see objectives such as "to give recommendations to inform future policy" or "to provide insight on how x could do y better in the future". In all cases, the evaluation is expected to provide recommendations to guide future action.

((w) Traditionally, and as discussed above, evaluation purpose has been either learning or accountability (or a mixture of the two). However, the Finnish Future Fund, Sitra, has taken a step further. In their new evaluation framework, **learning and accountability are accompanied by futures and foresight focus:** "...foresight knowledge will be used more explicitly and deliberately, and future-orientated nature embedded in surveying the operating environment, baseline situation analysis, preparing the theory of change, evaluation frameworks and the recommendations of the evaluations". A key point is that evaluation can support **transformative change** only if futures and foresight focus are part of the approach (Sitra, 2021).



((m)) The book "Visionary Evaluation for a Sustainable, Equitable Future" by Parsons et al. (2020) is another example which highlights the value-based role of evaluation and its role in building a "future of well-being for people, nature, and planet".



Evaluation purpose and objectives, as described in Terms of Reference (ToR), are often relatively future-oriented, at least implicitly. The gap between the past, present, and future becomes more apparent in the subsequent steps of the evaluation cycle. However, the ToR could include more explicit guidance on the role that futures-thinking is expected to play in the assignment. For example, a ToR might mention that the study will be forward-looking. However, the concept of "forward-looking" requires further clarification. It might refer to an emphasis on learning vs accountability (i.e. not dwelling on the "mistakes" of the past but trying to do better from now on). More concretely, it can refer to integrating specific foresight methods into the methodology.

Another interesting entry point to accountability is to ask **accountability to whom?** From the future perspective, it could refer to accountability to future generations. For example, climate justice includes the concept of *"intergenerational equity"*. The Fridays for Future is a global climate strike movement supporting this view (Fridays for Future, 2022). Another proponent of longtermism is William MacAskill with his recent book What We Owe the Future: A Million-Year View (MacAskill, 2022). Already the famous Brundtland report 'Our Common Future', which described the first official definition of sustainable development, referred to the future generations (WCED, 1987). Do you think the future focus should be brought as an equal element besides learning and accountability in evaluation? Why? What makes an evaluation transformative in your opinion?

2. Evaluation criteria

Evaluation is – by definition – a **judgement** of value. The values that the evaluation commissioner considers important are described in the evaluation criteria. The well-known criteria for development cooperation and policy proposed by the OECD DAC in 1991 are Relevance, Efficiency, Effectiveness, Impact, and Sustainability. The criteria were refined in 2019, and an additional criterion of Coherence was added (OECD DAC, 2019). These criteria are commonly used also in other sectors beyond development cooperation. Commissioners are, of course, free to use any criteria (or values) that they consider important.

Two OECD DAC criteria – Relevance and Sustainability – include some future elements. **Relevance** (i.e. is the intervention doing the right things) asks whether the intervention will "continue to do so if circumstances change". **Sustainability** allows to "determine if an intervention's benefits will <u>last</u> financially, economically, socially and environmentally". Interestingly, OECD DAC highlights that "Evaluators should also reflect on sustainability in relation to resilience and adaptation in dynamic and complex environments".







See the next section on formulating future-oriented evaluation questions under OECD DAC criteria.

((w)) There are also some attempts to evaluate the futures literacy of individuals, groups of people or organisations (Miller & Lianaki-Dedouli, 2015) and to develop related concepts and methods (APF, 2022). Sitra has also discussed what it entails to evaluate futures work. They discuss that assessment of impact would look at the production of futures knowledge and an increase in futures capabilities, and they can be observed at individual or organisational levels (Parkkonen, 2019). Although evaluations of organisations' foresight practices are relatively common, we understand that global good practices in evaluating futures literacy are not yet widespread.

The skill of **Futures Literacy** or the **Future Preparedness** of an organisation could be evaluation criteria that would capture the organisation's potential resilience to cope with unexpected changes. Futures literacy also includes envision-



towards the preferred future. Again, we observe a strong emphasis on values. Further, even if the evaluation did not extend its analysis to the time after the present moment, it is worth understanding whether any futures thinking took place during implementation. It helps to appreciate why certain decisions were taken during implementation. Kind and Wessels (2021) propose **Prospectivität** (in German; it could be translated as **Prospectivity**) either as a separate criterion or as an element embedded in other criteria.

Coming back to **accountability**, there is a link with evaluation criteria, namely, **impact** and **sustainability**. A colleague brought up a point of view that underlined that accountability to future generations is embedded in ensuring that development interventions create societal impact. And, moreover, those results sustain themselves without financial and technical dependency from donors.

How could or should future orientation feature in the evaluation criteria? What would be useful future-oriented criteria? Do you know of evaluations that have used this type of criteria?

3. Evaluation questions

Formulating evaluation questions is no simple task. Adding a layer of future orientation might feel overwhelming, but let's try it. First, let's do a quick recap; the main evaluation question types are descriptive (what is), normative (what is vs what should be), and cause-and-effect questions (how and to what extent the intervention contributed to solving the problem). Future orientation can be mainstreamed in all question types.

In a recent evaluation of Finland's water diplomacy (de Man et al., 2021), one of the evaluation questions was: "What is the current and **longer-term ambition** of the MFA and other Finnish key actors in the area of water diplomacy?" The question looks for specific answers about the future motivation of Finnish water diplomacy stakeholders in the current and future context. The team used an adapted Futures Frequency Workshop method to answer this (and other) evaluation question. We will come back to it in later sections.

At a minimum, evaluation questions under Relevance and Sustainability should be more explicit about how far in the future the benefits should last (e.g. one year, five years, or ten years?). For example:



Evaluation questions of relevance

Question type*	"Classic" formulation	More future-oriented formulation
Descriptive	Who received what services?	How are the needs of the stakeholders likely to change in the next ten years?
Normative	To what extent did the intervention respond to beneficiaries' needs as defined in the programme documents?	To what extent is the programme geared to ensur- ing that it meets the needs of the stakeholders in the next ten years, even if circumstances change? **

* Cause-and-effect questions are more applicable to the criteria of Effectiveness and Impact

** It is difficult to formulate a future-oriented normative question if there is no requirement for future orientation in the first place. However, the OECD DAC Relevance criterion includes the aspect of *"continue to do so if circumstances change"* which could be used as a benchmark in this case.

Evaluation questions of sustainability

Question type*	"Classic" formulation	More future-oriented formulation
Descriptive	What aspects of sustainability are con- sidered in the intervention (e.g. financial sustainability or technical capacity)?	How many years into the future are the benefits expected to last?
Normative	To what extent has the intervention set up the required systems to ensure financial, economic, social, environmental and institutional sustainability?	To what extent are the risks identified by the programme likely to affect the sustainability of the results in the next ten years?

^{*} Here again, it is difficult to formulate a future-oriented normative question if there is no requirement of future orientation in the intervention in the first place. However, the OECD DAC criterion of Sustainability includes aspects that have a direct link with future orientation. For example, the main definition is about "the extent to which the net benefits of the intervention continue or are likely to continue". The guidance also differentiates between "actual sustainability (i.e. the continuation of net benefits created by the intervention that are already evident) and prospective sustainability (i.e. the net benefits for key stakeholders that are likely to continue into the future)". Further, the guidance highlights the importance of examining the risks to sustainability, e.g. "the extent to which there are identifiable or foreseeable positive or negative contextual factors that may influence the durability of the intervention's results".

Evaluation questions of coherence

Question type*	"Classic" formulation	More future-oriented formulation
Descriptive	Does the programme create duplication of efforts internally or externally? What are other internal and external pro- grammes doing that have synergies with this intervention?	If other (internal and/or external) programmes fail in the next five years, can this programme succeed? What are other internal and external programmes doing and planning to do in the next 5-10 years that have synergies with this intervention?
Normative	Has the programme created synergies with other internal and external interventions as described in the programme document?	Not applicable.*
Cause-and- effect	To what extent is this programme likely to contribute to the poverty reduction of the population x at the end of the inter- vention if other (internal and/or external) programmes fail to deliver their services?	To what extent is this programme likely to contrib- ute to the poverty reduction of the population x if other (internal and/or external) programmes fail to deliver their services in the next five years?

Here again, it is difficult to formulate a future-oriented normative question if there is no requirement of future orientation in the intervention in the first place. There is also no built-in future dimension in the coherence criterion.
Going further in the discussion of how evaluation should take into account complexity, a coherence question could ask "How does the programme interact with other internal and external interventions" (normative)? The question could be modified to a cause-and-effect model, and the time aspect could also be added as above or in the form of sub-questions.



Evaluation questions of effectiveness and impact

When the time aspect is added to the evaluation questions under the criteria of effectiveness and impact, they become essentially questions of sustainability (namely, *prospective sustainability*). Therefore, we do not see the need to bend effectiveness and impact questions into a future-oriented format artificially. Here, we assume that the if an evaluation assesses the intervention's effectiveness and impact, sustainability will also be included as a criterion in the study.

What future-oriented questions would you like to include in the evaluation ToR? How would you strengthen the future orientation of the criteria of Relevance and Sustainability in the evaluation questions? Do you know of any evaluations that have included a prospective assessment of sustainability?

4. Evaluation scope

Evaluation scope refers to the time span, the number of interventions, or geographical areas that are included in the assignment. Evaluation ToR typically set the scope to 3-10 years in the past.

((v)) We do not have an example of an evaluation that would have established the scope to the past and the future. However, the *Evaluation of the Finnish Development Policy Influencing in the European Union* by Mackie et al. (2022) outlined significant near-future milestones relevant to influencing the EU's development cooperation (see Figure 1 below).

This relatively simple timeline allowed placing the evaluation conclusions and recommendations in a more concrete context than what would have been possible without it.

Figure 2: Some significant near-future milestones relevant to influencing the EU's development cooperation. Source: Mackie et al. (2022)









A low-hanging fruit to mainstream future orientation in evaluation would be to define the evaluation scope ten years to the future (in addition to the scope covering past years). In its simplest form, an assessment of future events could encompass an analysis of existing information about the future strategies of relevant organisations that interact with the intervention. It could also cover a compilation of the milestones that we will most likely occur. For example, we know that the Agenda 2030 is set to expire in 2030. This analysis can be included in the evaluation report's context chapter, in a separate annex, or even as a separate evaluation question.

Have you ever proposed to an evaluation commissioner to include an analysis of future events? Have you ever encountered an evaluation where the scope covered the past, present, and future? How many years to the past and future did it extend? Futurists typically set the future to ten years and after; what opportunities and challenges do you see here?

5. Approaches and research methods

Theory-based evaluation is one of the most established practices in development evaluation – at least, in the broad definition of the term. Hence, we focus on that approach in this blog. Similarly, there is often an implicit expectation that the exercise includes a strong utilisation focus.

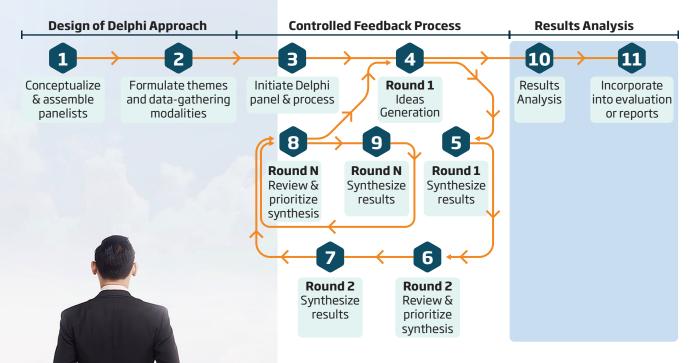
A theory of change can serve two primary purposes. First, it can be used as the evaluation framework to test whether the hypothesis holds ground or not (i.e. testing the assumptions behind the theory). Second, a theory of change can become a product of the evaluation. In other words, the theory of change might have to be reconstructed retrospectively before the evaluation starts; in the end, the revised theory can communicate what happened in reality. For an example of such a case, see Greenhalg et al. (2019).

The problem with theories of change is that they are **based on** the best-case scenario without considering the external factors that might impact the intervention. Theories of change also rarely define the period in which the change should occur.

A comprehensive evaluation of the World Bank Group's ((4)) renewable energy portfolio included a separate study using a **Delphi analysis** (IEG, 2022). The method is well known in foresight and consists of forecasting future events through an anonymous group communication exercise. The method differs from a regular survey by focusing on the question(s) "what could or should be". The purpose of the Delphi process in this context was to consult international experts on renewable energy to "help identify and prioritise emerging



renewable energy opportunities and challenges, and to **establish a future scenario** against which the Bank Group's capacity and position to influence can be evaluated". As the IEG report explains, the technique helps to "(i) shed light on alternatives; (ii) correlate expert insight on a specific subject; (iii) provide background information for decision-making; and (iv) reveal consensus in expert opinion. The method was applied alongside structured literature review, portfolio review and analysis, comparative case studies, and semi-structured interviews. This evaluation is, so far, the only one that includes a major foresight component that we know of.



((w)) The ParEvo tool developed by Rick Davies is a web-assisted programme for building future (or past) scenarios in a participatory manner (Davies, 2022). It has been used in evaluations, and as described by Davies "When used to look forward ParEvo can be seen as a form of participatory exploration of alternate futures. When used to look back it can be seen as a form of participatory public history". The website includes plenty of information on its applications.

The evaluation "Water as an Entry Point for Peace Mediation" commissioned by the MFA of Finland (de Man et al., 2021) applied a concise and adapted version of a **Futures Frequency Workshop**² in the data collection phase of the assignment. Members of the Reference Group participated in a session which discussed the different levels of ambition for Finland's water diplomacy work in the future. Based on the workshop and other information, the evaluation team structured the recommendations section to correspond to the different levels of ambition.

² Futures Frequency workshop, developed by Sitra, is a similar workshop format to the Futures Literacy Laboratory by Unesco. For more information, see <u>https://www.sitra.fi/en/projects/futures-frequency/.</u>

Figure 3: The Delphi process. Source: IEG (2022)



Here, we propose a few methods that evaluators could adapt relatively intuitively without significantly modifying the overall evaluation process. A "standard" theory-based evaluation could "easily" be combined with foresight methods, for exam-

ple, in the following ways:

- The context analysis could be powered by hori**zon scanning.** A "classic" context analysis typically outlines donor and partner country's relevant policies as a "mere" background chapter to the report. Quite often, it does not (unfortunately) play a key role in the overall analysis. Combined with horizon scanning, the analytical opportunities become more attractive. The primary purpose of horizon scanning is to identify possible discontinuities, emerging issues, and other signals of change relevant to the intervention or organisation, as Kuosa from the Futures Platform (2022) explained. The context analysis could also be separated into its own evaluation question (e.g. What future external factors might affect the intervention in the next x-xx years?). At a minimum, this analysis could look at what other actors in the field have identified as future trends and risks.
- The (re)construction and testing of the **theory of** change could include scenario building. In short, several possible pathways are developed instead of only one (the desired one). Also, the role of external factors is included more systematically. This approach would strengthen risk management and resilience when adverse situations (inevitably) occur.
- The circular **future wheel** is another method that allows more flexibility and embracing of complexity compared to the theory of change, which relies on a linear representation of the issue. It allows the creation of chains of "if, then" statements in the form of first-, second-, and third-order consequences.
- Backcasting examines a vision of a desirable or undesirable future and works backwards to identify the events that lead to it. In the case of undesirable futures, the methods help recognise factors that must be avoided to prevent that future from materialising.
- Wind-tunnelling is a method of putting an existing strategy to the test to assess how it might perform in different futures.
- Future Headlines/Cover page/Day in life and Artefacts from the Future are methods that present a future situation as if it was happening today.
- The internet hosts plenty of guidebooks on foresight methods that range from qualitative to quantitative, adapted to different time windows and levels of collaboration and co-creation. Some examples include the repository of futures and foresight toolkits

maintained by the Observatory of Public Sector Innovation (OPSI) (OPSI, 2022), the Playbook for Strategic Foresight and Innovation (Carleton et al., 2013), and Futures - Open to Variety: A Manual for the Wise Use of the Later-than-now by Bergheim (2021). In addition, the Foresight Manual by the United Nations Development Programme (UNDP) (UNDP, 2018) is explicitly adapted to the needs of development actors.

Do you know of evaluations that have used foresight or future-oriented methods? Which foresight methods would you like to test and why?

6. Data collection methods

Even if the evaluation would follow a fully "traditional" design without any future-oriented criteria or questions, there are still possibilities to mainstream futures thinking in the research process.

The Final Evaluation of Water and Sanitation ((p)) Programme for Small Towns in Vietnam, Phases I, II and III commissioned by the MFA of Finland (Greenhalg et al., 2019) is a "classic" ex-post evaluation in the sense that it did not include any specific future considerations as described in this discussion paper. However, the evaluation team probed the informants about when they expected the outputs, outcomes, and impacts to materialise. This aspect was not explicitly expressed in the programme documents and results frameworks. It turned out that stakeholders had varying views on this matter. As a result, the feedback helped the evaluation team to reconstruct the theory of change and discuss related lessons learnt.

A good start is being specific about what time window is considered. Is the evaluation considering a period only until the end of the programme? Or five or ten years into the future?

Possibilities at the data collection stage include:

- Asking informants in personal interviews or focus group discussions about the intervention's future. This can take the form of a SWOT analysis but is oriented to the years to come in addition to the present moment. E.g. what do you think are the most significant opportunities and threats to x in the x-x years to come?
- In an outcome harvesting and most significant change data collection process, the questionnaire could include a guestion about how the stakeholder thinks the outcome will evolve in the next year or further.



- Suppose the intervention's results framework or theory of change does not include speculations about the time needed to achieve the expected outcomes. In that case, the interviewee could probe this aspect and test whether programme stakeholders have the same understanding of the process.
- Asking final beneficiaries about their desired futures and how they think the future will probably turn out for them (e.g. about their professional life in a vocational education programme). These reflections allow assessing the programme's future relevance (the likelihood of the programme responding to the needs of the stakeholders also in the future).

In what ways do you think that future orientation can be included in the data collection phase in an evaluation which does not consider future-related aspects in the terms of reference or the initial evaluation design?

7. Evaluation conclusions section

The function of the conclusions section is to render a judgement about the intervention's successes and failures. Some interventions achieve or exceed their expected results. Other interventions remain short of their objectives. The evaluators should judge in the conclusions section whether the project/programme failed or whether the lack of achievements was due to external factors beyond the control of the implementers. Therefore, a failed project can still be judged a success because it was worth trying. If the conclusions are mirrored only against the past, an opportunity to reflect on the future context (which will - inevitably - be different from the previous) is lost.

The *Evaluation of the Finnish Development* ((y)) *Policy Influencing in the European Union* by Mackie et al. (2022) mentioned earlier (see the section on evaluation scope and Figure 1) built on the analysis of past implementation and future milestones to draw conclusions. Several factors combined together demonstrated weaknesses in the institution's forward-planning practices.

If the evaluation includes a future-oriented criterion, e.g., the institution's future preparedness, the conclusions section will automatically consider this aspect. However, if that is not the case, the context chapter

can still set the stage for tapping on the analytical potential of this part of an evaluation.

What is, in your opinion, the primary function of the conclusions chapter? Do you think that it is used to its maximal potential in evaluations? For example, how could it include future orientation?

8. Evaluation recommendations

The recommendations section is the most forward-looking part of evaluations. They should indicate who should do what and when and how high the priority is. Recommendations can be, for example, operational or strategic.

The evaluation of the World Bank Group's ((4)) renewable energy portfolio (that we have already mentioned earlier) future-proofed the recommendations using the Delphi method (see section Approaches and research methods). In September 2022, the Finnish Future Fund Sitra, adopted the same approach in the evaluation of the institution's Sustainability Solutions theme. The assignment is in the tendering phase at the moment of writing this paper; the results are expected in Spring 2023. The evaluation team is asked to produce short and longterm recommendations. Here, the Delphi method provides a solid framework for developing long-term strategic and thematic views on the way forward.

The report From Reactivity to Resilience - Assess-((p)) ment of the Response of Finnish Development Policy and Cooperation to the COVID-19 Pandemic by Laaksonen et al. (2022) recommended the MFA of Finland strengthen the crisis resilience of the Ministry's human resources by developing and implementing a headquarters crisis preparedness plan. This is an example of an ex-post evaluation (although titled an assessment, it is based on a robust evaluative methodology) that demonstrates a degree of future sensitivity.



Recommendations could include views on strengthening the institution's futures literacy or resilience. They could also be more specific about the short, long, and medium-term time windows of the recommendations.



Should recommendations include a more specific future focus? What would it be?

NIRÁS

Final words

It is apparent that when the inclusion of futures thinking in evaluation has occurred, the adopted approach has been more in the sense of future-as-time and not as future-as-image. In other words, evaluations include some ponderations about what relatively certain events will happen in the next years to come or what we already know that is in organisations' calendars. However, the idea of questioning our assumptions about the future and re-imagining possible futures to increase the usefulness of evaluation is in its infancy.

Ideally, the commissioners should integrate elements of futures thinking into the assignment's terms of reference. However, suppose that is not the case. In that situation, there are many opportunities in the subsequent steps of the project cycle to bridge the gap. A few suggestions include (i) proposing additional evaluation criteria, (ii) highlighting the future dimension of existing criteria, including time considerations in the testing of theories of change, (iii) formulating alternative scenarios, (iv) probing informants about future desires/ concerns, and (v) formulating recommendations that strengthen the institution's future awareness or adopt a concrete future vision (e.g. by applying the Delphi method).

Introducing the concept of future-as-time can open doors for discussions with the evaluation commissioner and evaluation team members to promote more advanced uses of foresight in the assignments.

QUESTION: Do you have other suggestions or examples that we did not cover in this paper? Let us know!

Petra Mikkolainen, Senior Consultant

Petra's background is in the implementation and evaluation of development cooperation and policy in multiple sectors. Her current areas of interest include the interconnections between evaluation and foresight and futures literacy in supporting organisations' and companies' strategy work. In past years, she has taken courses on futures literacy and foresight with the Institute for the Future and the Finnish Future Fund Sitra. In December 2020, she acted as a core team member in implementing a Futures Literacy Laboratory on the Future of Evaluation in Society. The event was organised jointly by Unesco, the European Evaluation Society, and NIRAS International Consulting as part of Unesco's High-Level Futures Literacy Summit. You can contact Petra by email <u>petra.mikkolainen@niras.fi</u> or via LinkedIn <u>https://www.linkedin.com/in/petramikkolainen/</u>

November 2022





Petra Mikkolainen NIRAS International Consulting November 2022