APF FORESIGHT EVALUATION TASK FORCE REPORT

Building Field and Foresight Practitioner Evaluation Capacity

December 23, 2022

ASSOCIATION OF PROFESSIONAL FUTURISTS

Prepared by:

APF Foresight Evaluation Task Force Members


Annette Gardner


Anisah Abdullah


Stephen Aguilar-Millan


Eric Barela


Stefan Bergheim


Laurent Bontoux


Shermon Cruz


Rick Davies


Nancy Donovan


Alex Fergnani


Kwamou Eva Feukeu


Jay Gary


Louisa Kastner


Bill Lesieur


Jewlya Lynn


Riel Miller


Jean Paul Pinto


Simon Schmitz


David Smith


Laurie Smith


Roger Spitz


John Sweeney


Bruce Tonn


Katri Vataja

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Background

In January 2021 the Association of Professional Futurists launched the APF Foresight Evaluation Task Force, a 24-member group of futurists, foresight practitioners, foresight evaluators, and evaluators.[1] The overarching aim was to standardize the quality of foresight practice and support the achievement of foresight aims. As identified in the 2019 themed issue of the *World Futures Review* on foresight evaluation, despite progress in foresight evaluation thinking and practice, the field has been slow to embrace the systematic investigation of foresight work, challenging learning and improvement.[1,2]

During its 18-month tenure, the Task Force engaged in a thoughtful examination of the state of foresight evaluation and the ways that evaluation can address the challenges raised by foresight work such as the long-time horizon for detecting impact. The Task Force explored the challenges and opportunities for expanding foresight evaluation capacity, including surveying the APF membership on existing evaluation capacity. It used this information to guide the development of useful and accessible evaluation capacity-building resources. The Task Force assembled evaluation constructs and methods to tackle many of the challenges posed by foresight and strengthen the implementation and quality of foresight. **Figure 1** below summarizes the Task Force's model of the benefits of evaluation to foresight practice, supporting practitioner excellence, improved program outcomes, and field building.

Figure 1: Benefits of Evaluation in Foresight

[1]Gardner and Bishop. "Expanding Foresight Evaluation Capacity." World Futures Review (2019) Vol. 11 (4) 287-291.
[2]Van Der Duin and Van Der Steen, 'Looking back on looking forward.' Futures. 44 (2012) 415-419.

Task Force Work Group Accomplishments and Recommendations

Task Force participants were divided into four Work Groups, three of which drew on the findings from the APF survey on member evaluation capacity conducted by Work Group 1. The Groups developed specific resources which could be completed in a 12-month period: an online curated foresight evaluation bibliography; a foresight evaluation guide; and an online foresight evaluation toolkit. Each Work Group developed a set of Recommendations, many of which are action steps that could be undertaken by APF, as well as suggestions for refining Work Group products. Work Group accomplishments and recommendations are summarized below:

Work Group 1: APF Member Evaluation Capacity and Resource Survey

The findings from the Foresight Evaluation Capacity Survey administered to APF members in September 2021 confirmed the observation that many foresight practitioners do not evaluate and/or do not know how to evaluate. However, more foresight practitioners said they do or can evaluate their work than anticipated. Respondents also identified specific foresight activities where they would like evaluation support: alternative scenarios, scanning, and foresight trainings/workshop, guiding the development of the online foresight evaluation toolkit.

Work Group recommendations:

- Create a new APF group of evaluation 'enthusiasts' to inform APF planning and discuss evaluation capacity building at the field building and practitioner levels;
- Establish a smaller, standing APF group to provide guidance to APF members on evaluation, such as 1:1 mentoring, review of evaluation plans, recommendations on methods;
- Post a very straightforward piece on evaluation on the APF website that then leads to all of the Task Force's resources;
- Provide regular evaluation training at APF conferences and stand-alone events;
- Host Annual Webinar for 'Clients' on the benefits of evaluation;
- Support an actual evaluation project, using internal funds; and
- Coordinate an APF-wide evaluation research proposal to be submitted to a foundation or agency like NSF, e.g., evaluation of public sector foresight programs.

Work Group 2: Development of a Curated Foresight Evaluation Bibliography

Developing a curated foresight evaluation bibliography was started early on to assess the state of foresight evaluation, identify resources that could inform Task Force thinking and activities, and develop a resource that could be used in the near-term by APF members. Informed by the survey findings, the Work Group collected a robust set of resources—peer review articles and reports—and populated an online bibliography in Zotero, an accessible platform.

Work Group 2 recommendations:

- The online Zotero bibliography should be made available to APF members, allowing for sharing of PDF documents;
- Identification of a dedicated person who is responsible for the ongoing maintenance of the bibliography and ensuring that existing content remains relevant and new content is properly added and tagged.

Work Group 3: Foresight Evaluation Guide

The Group prepared three frameworks to help with the evaluation of foresight:

1. The challenges of evaluating foresight and practical actions to address them;
2. Different approaches to foresight evaluation for different foresight approaches; and
3. Questions that might be asked to evaluate foresight with different purposes that seek to achieve impact at different levels.

The primary users for these products are foresight practitioners who would like to know how to better evaluate their work to increase its impact and its quality. Secondary users might include: Commissioners of foresight; Customers of foresight; and Evaluators of foresight.

Work Group 3 Recommendations:

- Test and finalize the frameworks with users;
- Develop professionally designed versions of the proposed frameworks. Presentation of the frameworks is as important as the content; and
- Explore and understand the use of foresight in evaluation as this seems to be at the methodological cutting edge.

Work Group 4: Online Foresight Evaluation Toolkit

Based on the APF survey findings, Work Group 4 designed and developed an online evaluation toolkit. It drew on other online evaluation and foresight toolkits and designed a landing page that could be included on the BetterEvaluation website. Work Group members developed two subsections that provide guidance and resources for evaluating alternative scenarios and scanning.

Work Group 4 recommendations:

- Work with the BetterEvaluation website team to support inclusion and maintenance of the toolkit;
- APF should play a leadership role in building foresight practitioner evaluation capacity, including having a part-time dedicated staff person (or consultant) who coordinates/conducts the following tasks:
    - Learning from foresight evaluation case studies;
    - Provide technical assistance, e.g., 1:1 mentoring;
    - Networking e.g., sharing evaluation best practices; and
    - Field building. Support evaluation of specific foresight approaches/products, processes, e.g., surface more evaluation cases in a WFR themed issue.

Recommendations to APF

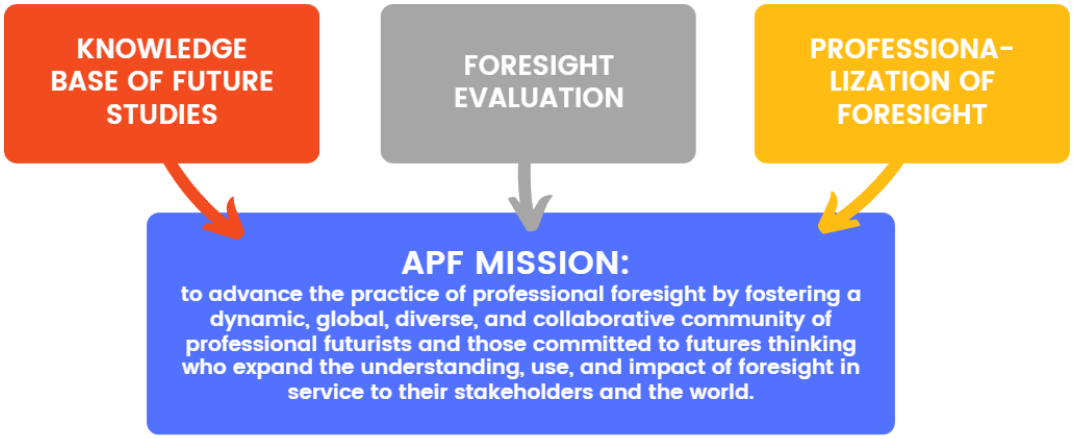
Based on Work Group activities and Task Force discussions, the Task Force developed two recommendations on how APF could play a leadership role in championing and advancing field and foresight practitioner evaluation capacity:

Recommendation No. 1: *APF embeds evaluation capacity building and practice in its organization structure, building an organization-wide culture of evaluation that supports the APF mission.*

APF could take a broader, larger approach and advance equally three core APF functions, specifically: *(see **Figure 2** below)*

- Knowledge Base of Futures Studies: which has been well-developed by Richard Slaughter and others as applied futures theory;
- Professionalization of Foresight: the APF Competency model, credentialing of APF members and;
- Foresight Evaluation: championed and built-out, e.g., Task Force products. Becoming a learning community of professionals to support individual competencies and organizational and field knowledge.

Figure 2: Achieving the APF Mission

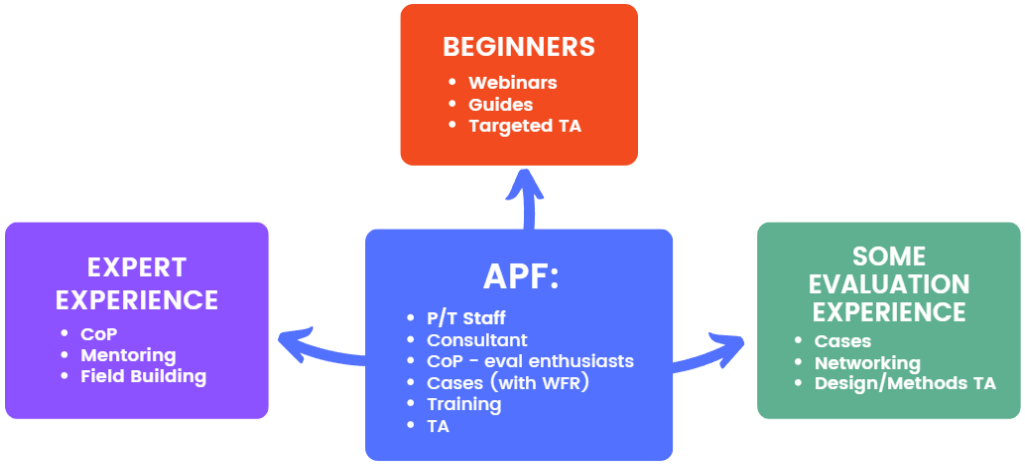


Recommendation No. 2: ***APF adopts a tiered model to develop and provide resources for foresight practitioners with different evaluation skill levels.***

A key learning from the APF Survey is that there is no 'one-size-fits-all' approach to providing resources. To address the diversity in foresight practitioner evaluation expertise, resources could be clustered by three skill levels of evaluation capacity. Evaluation resources would be targeted accordingly. (See **Figure 3** below)

As part of this model, APF could prioritize evaluation constructs and methods that improve the quality of foresight practice or process evaluation. For example, evaluation planning and theory of change (ToC) resources could be developed first, followed by impact evaluation resources

Figure 3: APF Foresight Evaluation Tiered Resource Model



Looking Ahead: Task Force Desired Outcomes

In summer 2022, at the end of the Task Force process, participants developed the following short-, medium-, and long-term outcomes they would like the APF Task Force and its products to achieve. Outcomes are categorized by four Targets: 1) Foresight practitioners, 2) Users of foresight, e.g., in strategic planning, 3) Commissioners of foresight, APF, and 4) Other, e.g., evaluators. These outcomes can be used for planning purposes as well as assessment of achievement. *(See **Figure 4** below)*

Figure 4: Foresight Task Force Short-, Medium-, and Long-term Outcomes

Target	Short-term: 1-2 years	Medium-term: 3-5 years	Long-term: 5+ years
Foresight practitioners	a. Usable evaluation findings to inform quality and effectiveness practice b. Increased awareness and use of evaluation tools/approaches	a. Increased evaluation competence b. Increase to 75% the APF members that evaluate their work, up from 30% in 2021 c. Improved quality and effectiveness of foresight	a. Increased credibility for foresight and evaluation b. New foresight evaluation frameworks and outcomes
Users of foresight, e.g., communities, decision-makers, organizations.	a. Increased user understanding of foresight b. Increased understanding of value added of foresight	a. Increased use of foresight in decision-making b. Increased use of foresight in strategic planning and preparedness	a. 'Foresight in all' public policy b. Prepared, nimble organizations, communities, government
Commissioners of foresight	a. Increased commissioner interest in evaluating foresight b. Increased understanding of value added of evaluation in foresight	a. increase in foresight Request for Proposals (RFPs) that have an eval requirement	a. Increase in evaluation expectations built into commission specs

Target	Short-term: 1-2 years	Medium-term: 3-5 years	Long-term: 5+ years
APF	a. Include eval in APF culture, e.g., competency model and/or model that underpins APF vision b. Test and finalize foresight evaluation frameworks c. Complete and put Task Force products into repositories that are broadly accessible	a. Certification/courses that include eval requirement. APF sponsoring a Coursera online certification class b. Participate/lead a multi-site evaluation of public sector foresight programs	
Others (evaluation field)	a. Leverage the marketing and PR the organizations these products are shared with e.g., BetterEvaluation	a. Research/Eval to examine foresight in decision-making—retrospective, multi-national study on orgs that use foresight and those that didn't	

Next Steps: Completing Work Group Products

Three Work Groups have developed products that require further development, review, and implementation, specifically:

1. Transfer the curated foresight bibliography to APF and promote use on a 'market testing' basis, e.g., document the number of users, resources viewed, documents downloaded.
2. User-test the Foresight frameworks and engage a designer to format and produce a Foresight Evaluation Guide.
3. Partner with BetterEvaluation to revise and host the online foresight evaluation toolkit. BetterEvaluation staff have indicated they are interested in hosting the toolkit in 2023.

While the Task Force has concluded its monthly meetings, individual Task Force members have agreed to be responsible for finalizing these products.

DETAILED REPORT

Background

A perennial concern of futurists and foresight practitioners is whether or not foresight work is having the desired impact. However, formal evaluation[3] of foresight work is rarely done. This is not due to lack of concern by the field. In 1981 in his aptly titled article, "How to Tell Good Work from Bad," Roy Amara argued that the field needed to develop criteria for judging the quality of its work if it was going to advance. He proposed general criteria for evaluating futures/foresight work, all of which are relevant today: conceptual explicitness, analytical clarity, and utilitarian objectives. And: does the work produce or guide action?[4] Despite 40+ years of scholarship on the benefits of evaluation and its application in actual foresight initiatives in the peer review literature,[5] the field has been very slow to assess the quality and impact of its work. In the 2000s we see increased application of evaluation to foresight and efforts to develop guidelines and models, such as the efforts by Georghiou and Keenan, Johnston, Rorhrbeck and Kum, Markarova and Sokolava, and Miles *(see **Figure 5**)*. In 2018, we see the same concerns by Fergnani and Chermack and others about whether futures studies (and by extension, foresight) can be viewed as credible if it's not willing to take a hard look at itself.[6]

On one hand the barriers are intellectual, e.g., whether thinking about the future is evaluable and/or how foresight evaluation fits with scientific inquiry and theory building and testing.[7] On the other hand, some futurists/foresight practitioners do not see the value in evaluating their work, which may reflect a lack of understanding of how and why evaluation is carried out. Increasingly, the evaluation arena is applying systems thinking and complex adaptive systems approaches to evaluating transformative initiatives, as well as adding foresight methods, such as scenario planning and the futures wheel to evaluation practice.

The tide may be turning. Foresight evaluation, through the slow accumulation of foresight evaluation thought, experiences, and resources, as well as advances in evaluation thinking and methods, is finally coming into its own, as evidenced by:

---

[3] Note: The Task Force's working definition of "evaluation" is: "...to provide (credible) answers to questions about a program (or activity) that will be useful…to inform action." (Rossi, Lipsey, and Freeman 2004.) 'Program' could refer to foresight trainings, process such as alternative scenarios, products such as movies, etc.

[4] Amara, R. How to Tell Good Work from Bad." *The Futurist*. 15(2) (April 1981), 63-71.

[5] Note: there are two themed issues on foresight evaluation: *Futures* 44 (2012) and *World Futures Review* 10, (2018)

[6] Fergnani, A. and Chermack. T. "The resistance to scientific theory in futures and foresight, and what to do about it." *Futures & Foresight Sci.* 2020.

[7] Rowland, NJ. And Spaniol, MJ. "On inquiry in futures and foresight science." Futures & Foresight Sci. 2020.

- Numerous publications in the peer review literature on foresight evaluation models and outcomes, as well as the findings from formal evaluations of foresight projects, including alternative scenarios, corporate foresight, and public sector foresight *(See **Figure 5**)*;
- Lessons learned from evaluating foresight and evaluation concepts and methods have been translated into foresight evaluation models, such as a section on assessing the impacts of participatory futures in the NESTA publication: *Our futures: by the people, for the people.*
- Evaluation as standard practice is being incorporated in mainstream foresight guides, such as a section on 'Evaluating Impact' in the 2020 RSA report, *A stitch in time? Realizing the value of futures and foresight*.[8]
- Foresight organizations are developing internal evaluation capacity, such as the The Finnish Innovation Fund Sitra's (Sitra) comprehensives process, outcomes, and impact evaluation approach.[9]
- Evaluators are beginning to learn about foresight and its role in supporting strategy and particular foresight methods that support evaluation practice.

Last, a culture of evaluation is forming, including the Association of Professional Futurists, the World Futures Studies Federation, Public Sector Foresight Network, and others which are contributing to the building of a foundation to develop field and individual evaluation capacity.

Figure 5: Published Foresight Evaluation Frameworks

- Georghiou and Keenan (2005) discuss evaluation strategies for national foresight activities and the importance of context;
- Johnston (2012) developed and tested a Foresight Impact Evaluation Schema that classifies four types of foresight impacts: Awareness raising, Informing, Enabling, and Influencing. Metrics for each category were developed;
- Makarova and Sokolova (2014) conducted a detailed literature review of evaluation of foresight evaluation projects and developed an evaluation model that builds on project evaluation but takes a tailored approach to evaluating foresight, such as using primarily qualitative methods;
- Miles (2011) proposed a dynamic approach to evaluation that focuses on the interactive nature of foresight in service terms and not just formal inputs and outputs;
- Piirainen, Gonzales, and Bragge (2012) developed an input-process-output framework to be used in designing an evaluation, enhancing impact and generating credible evidence;
- Poteralska and Sacio-Szymanska (2014) examined national and transnational foresight projects and provided a review of systemic foresight evaluation frameworks and their application;
- Rohrbeck and Kum (2017) developed a model to assess a firm's future preparedness and impact on firm performance;
- Steen and Twist (2012) developed an impact evaluation framework that consists of five elements and provide suggestions for evaluating the impact of futures studies in the public sphere to help bridge the gap between foresight and policy.

[8] Shallow, A. et al. A stitch in time? Realizing the value of futures and foresight. RSA. October 2020.

[9] Link to the Sitra approach: https://www.sitra.fi/app/uploads/2022/04/sitra-evaluation_framework_december_2021-006.pdf

Task Force Purpose, Aims

Recognizing the need for strengthened foresight practitioner evaluation capacity as well as increased assessment of foresight projects and activities, the APF Foresight Evaluation Task Force was launched in early 2021. The 24-member Task Force was intended to support increased evaluation capacity through sharing of evaluation resources, strategies, designs, and methods, and provide suggestions for appropriate foresight outcomes and indicators to guide evaluation design. As described in the Task Force Charter in ***Appendix A***, the overarching aim is to standardize quality of foresight practice, and support achievement of foresight aims.

Task Force Structure and Activities

Reporting to APF, The Task Force met monthly from February 2021 until July 2022 (18 virtual meetings). The Task Force was comprised primarily (but not exclusively) of 24 individuals from APF and included representatives from the World Futures Studies Federation, the American Evaluation Association, the UNESCO Foresight Department, the Public Sector Foresight Network, and others.

The Task Force had standing 1 ½ hour-long monthly Zoom Meetings. Typically, participants engaged in a 30–60-minute discussion of an issue or topic selected by the Task Force Chair, Annette Gardner, PhD, such as the state of foresight evaluation practice or a technical discussion on evaluation design. Task Force discussion items included:

1. Determine the challenges to evaluating foresight or 'evaluability,' such as resource constraints, a long-time horizon, etc.,
2. Clarify the state of evaluation practice by futurists, e.g., survey APF members to assess their work, challenges encountered, methods used, etc.,
3. Identify evaluation designs, outcomes, methods, and tools that have the potential to further inform practitioner excellence in foresight, and
4. Develop an 'evaluation capacity building' (ECB) model for expanding APF member evaluation expertise.

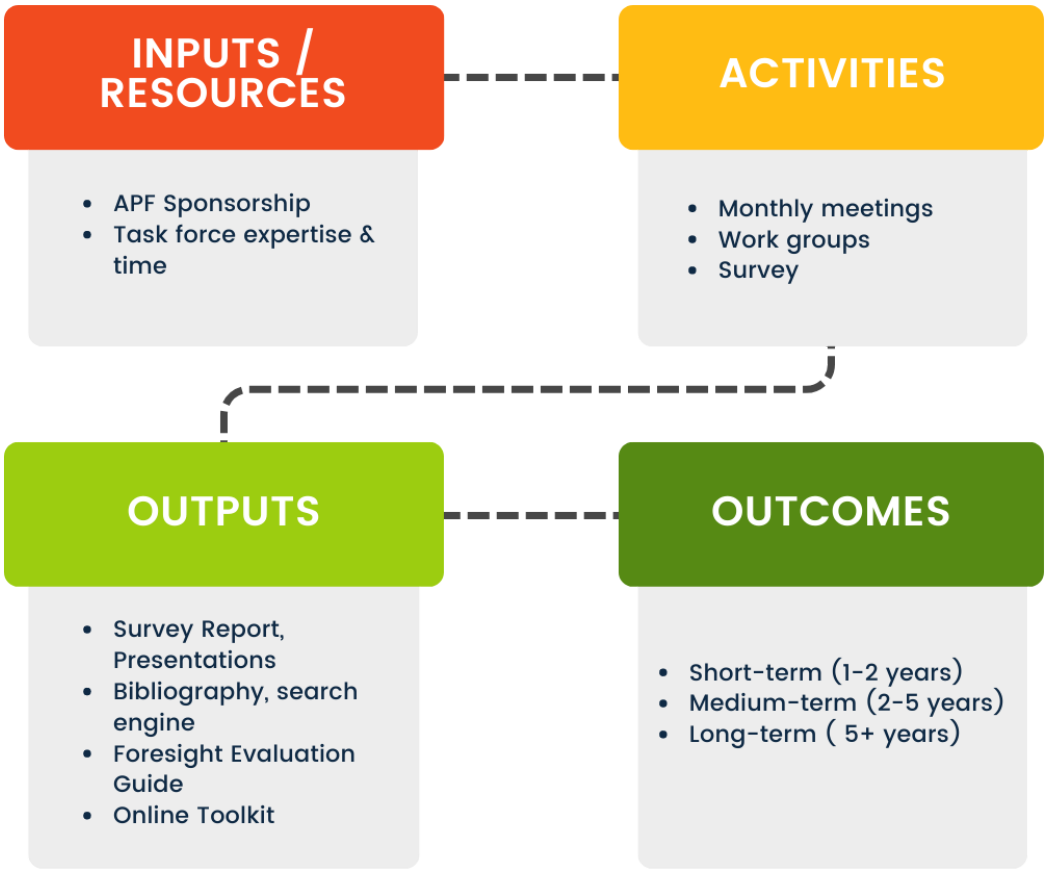
Second, four Working Groups were launched to accomplish specific tasks:

- Work Group 1: Develop and administer a survey to APF members in fall 2021 to assess member evaluation capacity and desired resources;
- Work Group 2: Develop an accessible, curated foresight evaluation bibliography;
- Work Group 3: Develop a foresight evaluation guide; and
- Work Group 4: Develop an online foresight evaluation toolkit.

Resources, Leadership, and Reporting

Task Force Chair, Annette Gardner, PhD, was responsible for chairing the virtual meetings on Zoom. A futurist who is also trained in evaluation, she developed the Meeting Agendas, and sent out resources before and after the Meetings. She provided evaluation capacity building resources to the Task Force and APF membership via the APF website. Jay Gary, PhD, supported Task Force communications, posting Meeting recordings and descriptions on the APF Website, under the Evaluation section. See ***Figure 6*** for a model of the Task Force resources, activities, and outputs.

Figure 6: APF Foresight Task Force Process Model



Work Group Reports and Recommendations

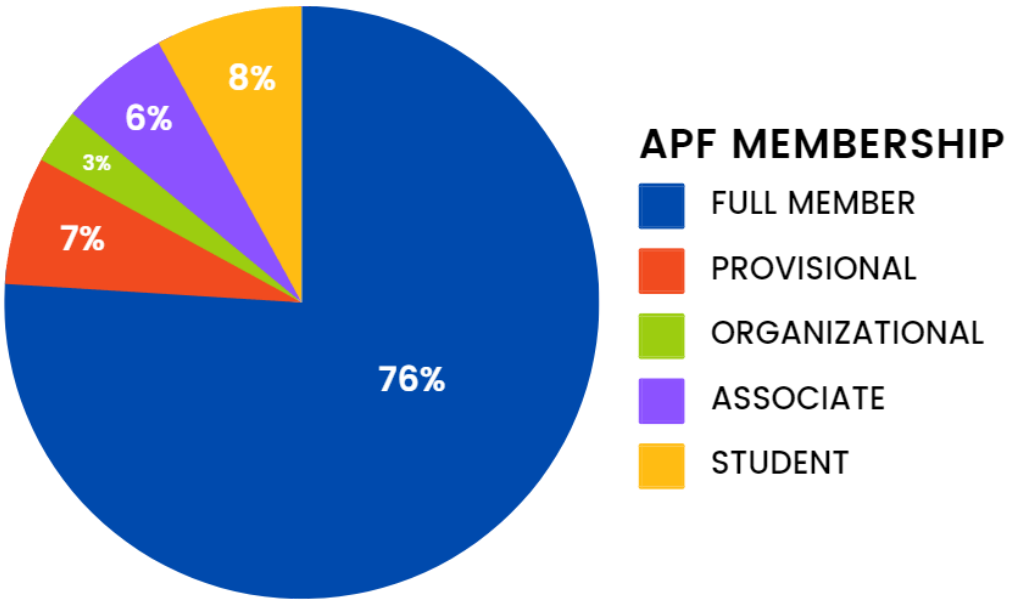
Work Group 1: APF Member Survey–Foresight Evaluation Capacity

Work Group 1 consisted of 3 members of the Task Force:

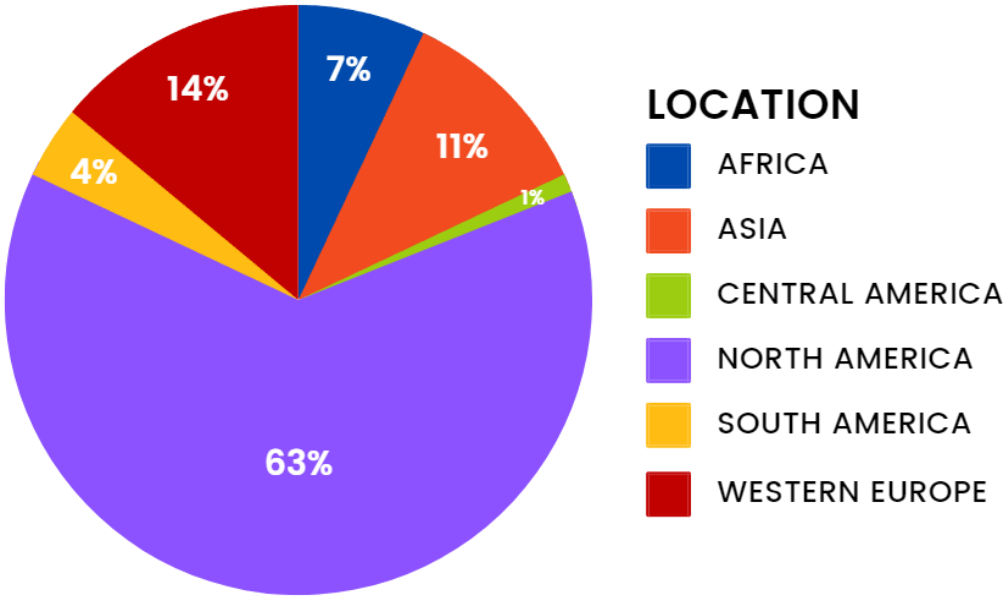
- Bruce Tonn
- Bill Lesieur
- Annette Gardner

The work group designed the Foresight Evaluation Capacity Survey administered to APF members in 2021 and provided the results and insights back to the other work groups to support their respective activities and prioritization of areas of need. Survey questions focused on: respondent evaluation expertise and practice, challenges to evaluating foresight, and desired support and resources that APF could provide on the subject of evaluation of foresight *(See **Appendix B**)*. The survey was administered electronically to the full APF membership (500 members) during the August-September 2021 time period. One reminder was sent to non-respondents. One hundred seventeen individuals responded, which is approximately 20 percent of the APF membership. Survey findings are described below:

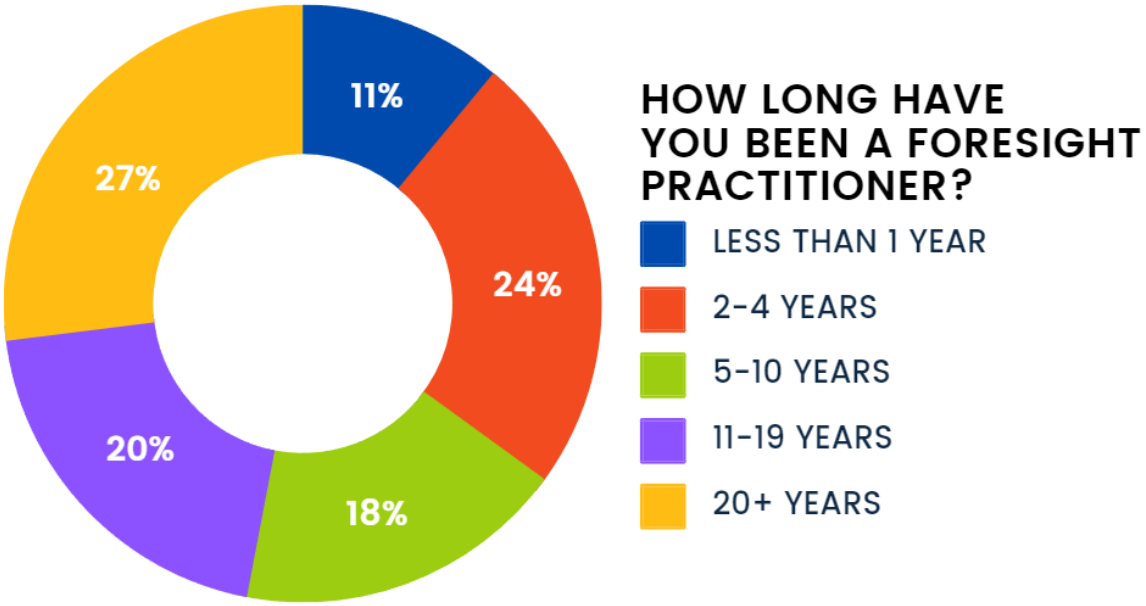
Respondents were mostly Full members, but other categories were represented too.



Respondents tended to be from the Northern Hemisphere.



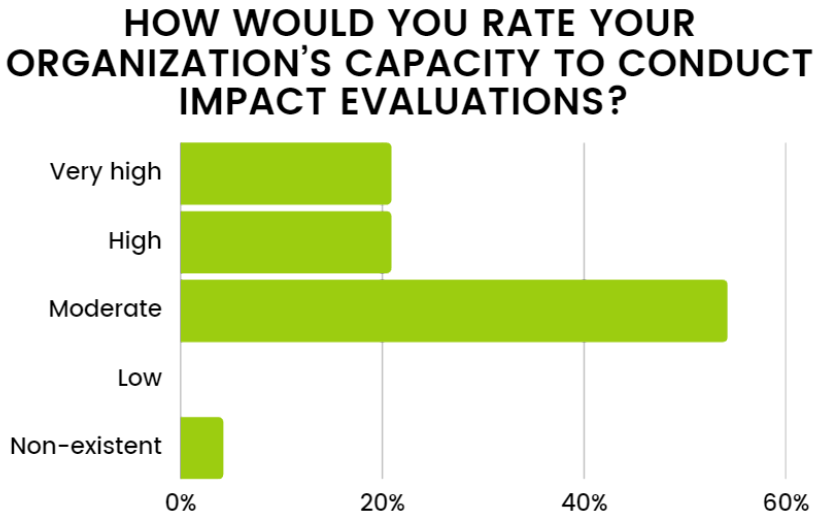
Respondents represented a range of practitioners—from less than 1 year to 20+ years of experience.

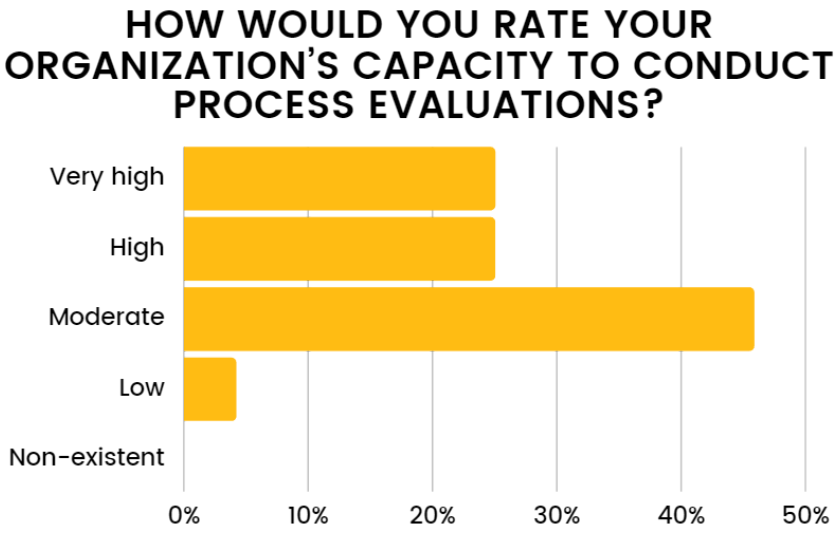


Respondents (n = 107) indicated there are many reasons for evaluating, with many emphasizing learning to improve foresight (process evaluation) and assessing foresight effectiveness or outcomes evaluation.

Overall, the findings speak to moderate evaluation capacity amongst APF members though a low use of evaluation.

Also, respondents report being equally adept at both process and impact evaluations.

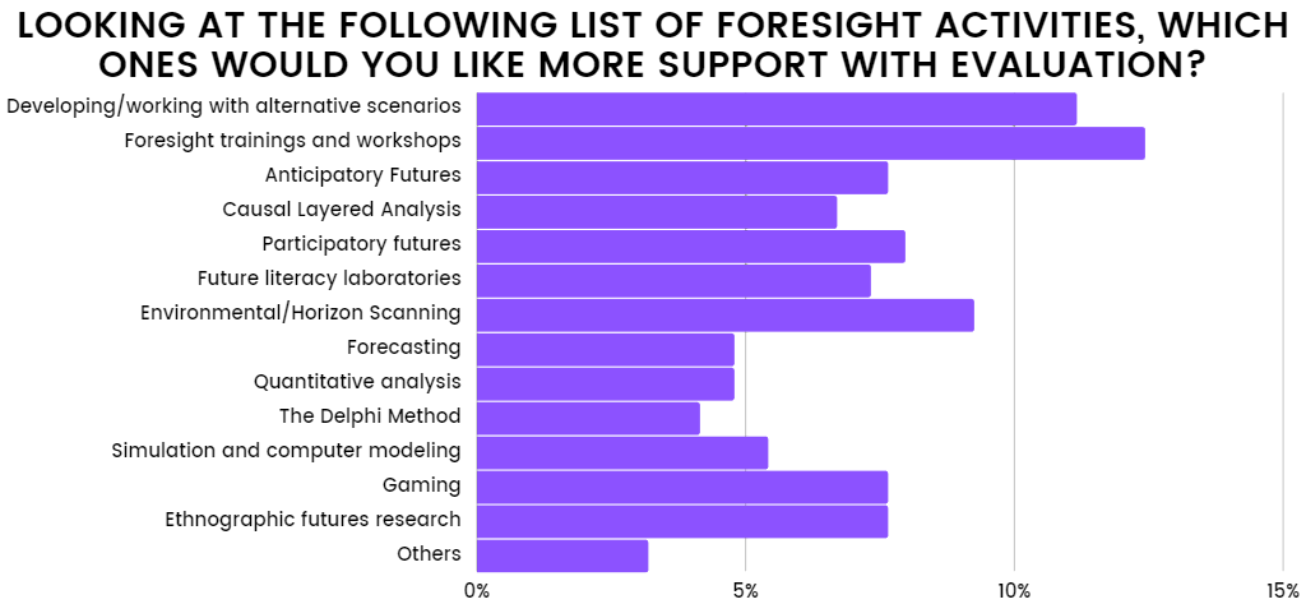


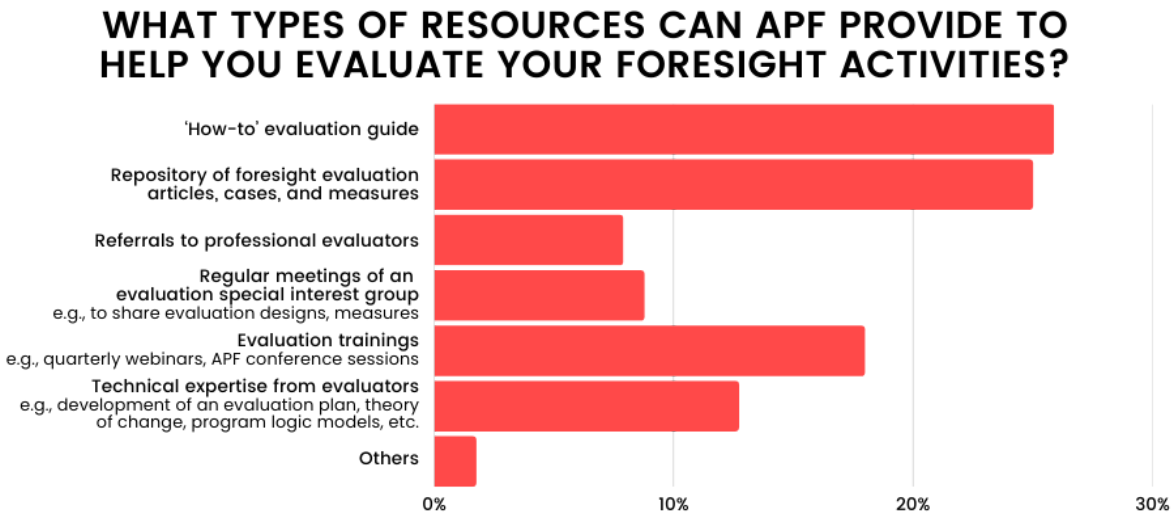


Respondents were asked about the biggest challenges to evaluation of foresight. Approximately 35% of responses cited *lack of resources*, such as limited staff time and insufficient financial resources. Almost a third of the responses cited more *technical issues*, such as lack of knowledge amongst staff and about data collection and analysis issues. The remaining responses, also about one-third of the total, referenced *lack of strategic support for evaluation*. For example, clients may not be interested in evaluation and there are barriers to follow-up and use of evaluation results.

Resources (35%)	
Limited staff time	13%
Insufficient financial resources	10%
Insufficient support from organizational leadership	8%
Insufficient support from organizational staff	5%
Knowledge (12%)	
Limited staff evaluation knowledge, skills, and/or tools	7%
Not knowing where / how to get started with evaluation	5%
Process & Data (19%)	
Difficulty in specifying variables to measure impacts	7%
Difficulty in collecting the requisite quantitative and qualitative data	5%
Difficulty in developing rigorous designs, e.g., valid, generalizable	5%
Difficulty in analyzing data	3%
Purpose, Use, & Stakeholder Expectations (31%)	
Lack of follow-up and use of evaluation findings, e.g. to inform strategy	10%
Clients are not interested in program achievements and longer-term impacts of activities	10%
Lack of appropriate outcomes, indicators, and methods that fit my organization's work	6%
Managing funder expectations and/or reporting requirements	4%
Difficulty in communicating results	2%

The results on specific areas for evaluation support and resources desired suggest there is a role for APF to assist members and their organizations improve their knowledge about evaluation and evaluation-related analytical abilities. Respondents indicated that how-to-do-evaluation guides and a repository of papers and cases, particularly in the areas of workshops and scenario development, would be of most value. At the field level, APF could take a leadership role in increasing commitment to the improvement of the discipline of foresight that may not materialize in the immediate context of any one project.





Work Group recommendations include:

- Create new APF group of evaluation 'enthusiasts' to inform APF planning and discuss evaluation capacity building at the field building and practitioner levels;
- Establish smaller, standing APF group to provide guidance to APF members on evaluation, such as 1:1 mentoring, review of evaluation plans, recommendations on methods;
- Post a very straightforward piece on evaluation on APF website that then leads to all of the Task Force's resources;
- Provide regular evaluation training at APF conferences and stand-alone events;
- Host Annual Webinar for 'Clients' on the benefits of evaluation;
- Support an actual evaluation project, using internal funds; and
- Coordinate an APF-wide evaluation research proposal to be submitted to a foundation or agency like NSF, e.g., evaluation of public sector foresight programs.

Work Group 2: Draft Curated Bibliography and Analyses

Work Group 2 consisted of 5 members of the Task Force:

- Rick Davies
- Alex Fergnani
- Annette Gardner
- David Smith
- John A. Sweeney

Activities

The work group's activities were of three kinds:

1. Searching for relevant papers to include in a bibliography
2. Identifying an appropriate software package to store the bibliography
3. Filtering the found papers for relevance for the task
4. Structuring the filtered papers according to a category scheme that would be both meaningful and usable to members of the APF.

Searches

The following sources were explored:

1. Papers known to the members of Work Group 2, and other Task Force group members. These were relatively small in number but provided an initial seed giving the Task Force an idea of what they were aware of. Subsequent search activity showed that these papers were a very small subset of the total number of papers that were available on the subject of foresight evaluation.
2. Searches via Google. This kind of search had a very wide scope potentially taking into account not only published papers but also some of the gray literature that had found its way onto the web, plus content from organizations' websites as well as personal blog postings.
3. Searches via Google Scholar. This kind of search had a much narrower focus but produced content which was more visibly relevant.
4. Searches via specialized bibliographic search sites such as Semantic Scholar. '*Semantic Scholar uses groundbreaking AI and engineering to understand the semantics of scientific literature to help Scholars discover relevant research*".
5. Searches via existing bibliographic networks. A number of software packages are now available which visually display all the papers referenced in one or more seed documents of your own choice. Papers which are referenced by more than one of those seed documents are likely to be of particular interest. LitMaps was used in the last stage of this task force's work, to explore references cited by papers already identified as relevant.

Search terms

The choice of which words to enter into a search engine of any of the above kind will obviously make a huge difference to what is found. In our explorations our use of search terms was not very systematic. At different points in time they included: *foresight evaluation, scenario evaluation, alternative futures evaluation*, etc. The most common search terms involved combinations of a reference to a foresight synonym and the word "evaluation." We anticipate that members of the APF will make their own choice of relevant search terms, and don't see any value in recommending a specific set of terms here. Consistent and careful use of search terms will be most important where the user is trying to do some scientific study or literature review of the field, where a comprehensive and transparent level of coverage of available documentation is important.

Filtering

The main reason why some filtering was needed was to define the boundaries between papers which were relevant to the subject of evaluation of foresight and those which were not. These excluded were:

- Papers about forecasting, based on quantitative analysis and simulations
- Papers which seek to review and or develop conceptual frameworks which can be used to think about foresight practice. These were excluded from our accumulating bibliography unless they made explicit reference to the relevance of that framework for some evaluation purpose.
- Literature reviews of foresight practice in a particular area, which were largely descriptive but sometimes have some evaluative content.

Caveat: Filtering out of papers which weren't sufficiently relevant was not done on a particularly systematic basis. There are likely to be some papers in the bibliography that now exists which should be removed.

Storage

The papers which were identified as relevant have been stored using a bibliographic database software package known as Zotero. This is a widely used package with good import and export facilities into other formats such as those used by Mendeley, another widely used package of the same kind. Their relative advantages are well described here. Zotero enables cloud as well as private storage of bibliographies.

Towards the end of the work group's activities another software package, known as LitMaps, was identified as potentially very useful. Zotero and other bibliographic collections can be imported into this app, which can then be used to generate a scatter plot showing how those papers are distributed over time (X axis) and by number of citations received (Y axis). LitMaps includes an AI assisted search for other relevant papers connected by citation links to those in the existing bibliography. It also allows collaborative annotation of the papers in the bibliography. Its full potential, and limitations are still being explored.

Structuring

While not hugely time-consuming this task was the most cognitively challenging for the task force members. We were looking for a conceptual framework that would help APF members search the accumulating bibliography for papers that might be of most use to them. At the beginning we settled for a single layer of five categories. For example: foresight evaluations, foresight evaluation frameworks, foresight

measures, foresight evaluation resources, and evaluation resources outside of foresight. Later on, we developed a nested category structure (i.e. a branching tree structure), which developed successively more detailed categories and subcategories, 18 in all. One or more example papers were provided at the end of each branch, exemplifying that particular kind of paper. Some papers were placed in more than one category. (***Figure 7***)

Figure 7: Bibliography Structure

Evaluations by type of foresight method	Delphi			8
	National foresight systems, e.g.,			8
	Online foresight platforms, e.g.,			2
	Scanning			4
	Alternative Scenarios, e.g., conceptual frameworks, cases,			7
Stages of foresight use	All stages			2
	Process	Process design		2
		Training		1
		Implementation		4
	Outcomes	Examples	Scenario readiness	1
			Individual learning outcomes	5
			Longer term benefits	6
Types of foresight users	Corporate			3
	National			4
	Public sector			1
	Student			2
Evaluation frameworks - generic/overall				13
				73

Caveat emptor: A quick look at these categories will prompt thoughts about the many apparently missing categories that come to mind. They are missing, but only for the time being, because no papers have been found which exemplify evaluation practice in that area.

This tree structure was initially made visible as a web page. Later on, David Smith subsequently developed an Excel version where rows described references and columns described categories and sub-categories, etc., enabling sorting of references by different levels of categories, or combinations thereof. The downside of these two experiments (the webpage and Excel version) is that they were not easily updatable, so their further development is not recommended. However, the nested category structure has since been incorporated into the structure of the online Zotero bibliography.

More recently we have given more attention to the use of tagging, a facility available in both Zotero and LitMaps whereby one or more tags such as #corporate, #delphi could be applied to a particular paper in the bibliography. Then searches could be made for any paper with any specific combination of tags which were relevant to the person making the search. Work is still underway to make all the categories mentioned in the previous two paragraphs searchable by tags.

Final products

Two versions of the bibliography are currently available:

- APF Evaluation Library Vs2 available via Zotero Online. Here the references are sorted into 18 categories and subcategories, as shown above
- APF Foresight Evaluation bibliography via LitMaps. Here an almost identical set of references are shown in a way that displays the relevant papers they reference, and other papers that have subsequently referenced them.

Recommendations

- The Zotero bibliography can be made available in private or public form. The public form is accessible by anyone in the world who has access to the link to the bibliography. But Zotero does not allow attachments such as PDF copies of papers to be shared in a public form of the bibliography. PDF papers can be attached to references in the private form of the bibliography. But private bibliographies require users to become approved members of the bibliography and have their own username and password;

- We recommend that the private version be made available to APF members, from within the APF website that is accessible to members only. This might require some coding by the managers of that website such that anyone who logs onto the website as a member is automatically treated as a member of the private Zotero bibliography;
- The Zotero bibliography will need some form of ongoing management. New papers may be identified and proposed, and periodic searches for new relevant papers would properly also be advisable. When this happens it is highly likely that the existing category structure will be seen to need some form of adjustment and/or expansion in categories;
- We recommend that at least one person is responsible for the bibliography. But that person is supported by a small 'study' group. That study group could identify and solicit papers which are potentially valuable and jointly read and discuss them in periodic meetings in a format similar to that used in book clubs. That group would have to be voluntary in membership and have a say in what papers it wanted to read. Its advice could then be used by the manager of the bibliography, who would update the contents of the bibliography and possibly the category structure, after each new paper was read and discussed; and
- Bibliography already uploaded into Litmaps could be made available to APF members without the need for any additional coding in order to ensure password-based access. Its use is likely to be bought limited because it is more technically demanding, but it is likely to be of value to the manager of the bibliography and any supporting study group.

Work Group 3: Draft Foresight Evaluation Guidelines

Work Group 3 consisted of four members of the Task Force:

- Stephen Aguillar-Millan
- Laurent Bonteux
- Laurie Smith
- Katri Vataja

What activities did the working group undertake?

The working group developed the frameworks over a series of meetings during 2021 and 2022. The initial meeting was to set the project brief and subsequent meetings involved iterating through various models of potential frameworks.

Products

Members prepared three frameworks to help with the evaluation of foresight (see below):

- The first describes some of the challenges of evaluating foresight and proposes practical actions that might be taken to surmount them.
- The second describes different approaches to foresight evaluation for different foresight approaches.
- The third describes questions that might be asked to evaluate foresight with different purposes that seek to achieve impact at different levels.

The primary users for this product will be foresight practitioners. A provisional use case for the primary audience might be: "As a foresight practitioner I would like to know how to better evaluate my work so that I can be more effective and increase the impact of foresight."

Secondary users might include:

- Commissioners of foresight
- Customers of foresight
- Evaluators of foresight

The frameworks are intended to help guide the thinking of busy foresight practitioners so are intended for quick, easy use.

The Work Group is not aware that frameworks exist that fulfill the same purpose. The frameworks do, however, draw on previous work such as that of Ramos et al (2019) and Sitra.

Framework 1: The challenges of evaluating foresight

The Framework below describes some of the challenges or potential questions/suspicions in evaluating foresight, a description of these challenges and potential solutions to tackle these challenges.

Challenge	Description	Potential solutions
Certainty	Decision makers and "customers" who desire evaluation often want certainty. Did a foresight process work as intended? Will the foresight process reduce uncertainty? Foresight helps users understand and recognize uncertainty better, thereby helping people deal with it better.	Using evaluation to better frame foresight projects that intrinsically acknowledges the inherent uncertainty associated with foresight methods. Educating decision makers / evaluators / customers that foresight explores uncertainty, so they do not expect too much certainty. Educating foresight practitioners of the need/desire for certainty among decision makers, evaluations, and customers. Using theory of change (ToC) as a tool to illustrate the purpose of foresight, its aims and assumptions of what they want to see happen as a result of the foresight. What are the relevant questions for evaluation in relation to the ToC?
Timing and salience of impact	The challenge of prediction. The impact of foresight is often long-term and can be something that didn't happen (e.g., the pandemics that were avoided). Foresight can be used for preparedness and prevention. As a result, foresight success will be achieved by adverse outcomes not happening, something difficult to ascertain.	Understanding that predictiveness is not a desired impact of foresight. Nor is it a prerequisite for evaluation. Interim / milestone measures that might show progress toward outcomes that might be further in the future. Proxy measures that can be measured more immediately than final outcome measures. Theories of changes that show the potential pathway to impact, including interdependencies of effects and outcomes, and assumptions in relation to the change and context, even if that is not reached any time soon. Process evaluation to understand whether methods were well implemented even if the outcomes are unknown. Note: process evaluation can be valuable in its own right rather than simply as something to use when outcome evaluation if not possible. Process evaluation complements outcome and impact evaluation, helping to understand what happens and why and what works in what kind of context. Without process evaluation, evaluation is a so-called "black box" and its results are difficult to use in development or scaling. Active exploration and evaluation of counterfactuals to

		consider what didn't happen, which is often neglected. Foresight is not for predicting the future but informing present choices/decisions. Thus, evaluation is needed to inform if and how foresight generates its outcomes, e.g. how foresight has informed decision-making or help to increase people's preparedness for the future.
Goals	Different kinds of foresight approaches have different goals, which are not always agreed upon. They are also many and can be difficult to define.	Clearer problem / outcome definitions so goals are explicit and understood by all involved in a foresight engagement. Clearer taxonomy of futures methods and the sorts of problems that they can solve. The use of evaluation approaches that aim to identify goals, such as goal-free evaluation and facilitation around goals are a good first step. Developmental evaluation and pre-program planning is another approach that aims to identify intended/unintended, desirable/undesirable, anticipated/unanticipated outcomes
Attribution	Attribution and contribution of outcome and impact from foresight is often difficult and the causal chain between intervention, outcome and impact can be unclear and long.	Theories of change that can illustrate the attribution or contribution of particular individuals or groups within a larger whole. Adapted Bradford Hill criteria to better understand and attribute causality when understanding and attributing the role of foresight in outcomes and longer-term impact. The use of systems thinking methods, e.g. systems maps, to better understand attribution and contribution by multiple actors to an outcome and a longer-term impact. Using evaluation methods that focus on understanding complexity and analyzing contribution instead of attribution; e.g. systems change approaches, contribution analysis, process tracing.
Craft	Foresight methods are a craft with lots of tacit knowledge and methods are mixed or adapted, most often tailored to the specific circumstances of a project.	Clearer taxonomy of futures approaches and methods, their characteristics and what they can achieve. Developing a framework to identify and understand archetypes of mixing foresight methods, their characteristics and their value. Understanding that evaluation is not something "designed" for only standardized or highly developed models/actions or simple/static contexts and 'best practices'. The object and context of evaluation is rather messy and complex in all relevant/important evaluations. Evaluation could be used as a tool to make tacit explicit, to facilitate learning by analyzing how and why the path has evolved during

		implementation, what is relevant, what works, what have been learned and what is scalable or transferable into other contexts.
Cost	Evaluation can be expensive, particularly at scale, and foresight projects often do not have substantial resources.	Assess the applicability and feasibility of experimental, semi-experimental, and non-experimental or descriptive evaluation approach. RCT is not ideal for complex, uncertain foresight activities and programs Evaluation does not always have to be expensive: 1) Prioritize what kind of evaluation knowledge is needed and utilized (and by whom). Is evaluation for internal needs and developmental purposes or more for accountability and external use? 2) Plan from the start what kind of knowledge could be produced during the project/process for the use of monitoring, evaluation and learning. Educating funders to require evaluation and build that into funding in an appropriate way that supports foresight in implementation and learning too. Developing evaluation as an investment for getting better/impactful projects in the future.

Framework 2: Simplified framework for selecting an evaluation approach

As noted above among the challenges, the foresight field consists of different kinds of approaches (and paradigms) and foresight can have many kinds of goals. How to choose an appropriate evaluation approach and methods for your foresight? What are the relevant questions that guide designing evaluation, meaning what kind of evaluation questions could be set and how to produce evaluation knowledge of them? The aim of Framework 2 is to help to identify a purposeful evaluation design and suitable methods based on the nature of foresight and help practitioners to start navigating existing evaluation repositories, such as the APF Foresight Evaluation Bibliography and BetterEvaluation online toolkit.

What is the goal and focus of the foresight activity?	**Support to (contingency) planning and strategy development**	**Visioning, development, and adoption of a shared vision**	**Transformation, futures literacy, foresight capacity, connecting futures thinking to making change**
What are the endpoints to evaluate?	Scenarios Drivers of change Milestones towards the future Applied analytics 'Systemicity' Contribution of foresight to a theory of change or other decision-making	Improvement of the coherence of action between the various 'silos' of an organization Long-term 'uptake' of the vision	Capacity
What kind of evaluation questions and criteria are typical/useful?	Methodological adequacy (fitness for purpose) How "futureproof" is the strategy? How alternative developments have been considered? Preparedness/agility/adaptiveness of decision making and strategy Quality of process, (robustness, inclusiveness, etc.)	What was the value of vision: as a process and a product? Quality of process (robustness, inclusiveness, etc.) Commitment Coherence of action	How futures thinking has been linked to action? Changes in individual's/ communities' futures thinking and capacity? Novelty in action? Decolonializing of futures? Diversity Inclusiveness
What kind of evaluation approach is suitable?		Qualitative case study Process evaluation Before and after assessment of coherence of action between 'silos'	Qualitative case study, stakeholder survey/ interviews, self-capacity measurement, ethnography, action research

How to use the framework?

1. The first task is to identify what is the purpose and focus of the foresight that will be evaluated?
2. What are the interests of the evaluation? What to evaluate and what is relevant? What are the criteria for success?
3. Based on the answers to the first two questions and on the constraints linked to the evaluation exercise, choose the methods that are best suited to generate the desired answers.

Background to Framework 2: How to evaluate different foresight practices? Different approaches to evaluation

The first point: *Foresight and evaluation as information practices have evolved side by side over the past decades, but the potential for cooperation has not yet been fully realized.*

On one hand, how could foresight be better evaluated? On the other hand, how could foresight and futures thinking make evaluation more future-oriented and therefore more useful, especially for *ex ante* impact assessments?

There are several definitions of foresight, but in general it is a reflection on how best to understand current changes, what they mean for alternative futures and how this can inform measures to achieve the desired future. In postnormal times, living with surprises and uncertainty, making precise and straightforward action plans with an emphasis on efficiency becomes both useless and impossible; focusing on continuity, business-as-usual and "return to the normal" not only is not a useful way to perceive the world but prevents from being able to seize opportunities to create a better future. Postnormal times thus challenge everyday life and make foresight and the practices of futures work more important. Today, what is needed above all is (i) a diverse and inclusive debate to understand the world and imagine desirable and achievable futures, (ii) the ability to tackle unexpected phenomena (resilience), and (iii) building the capacity to adapt and find ways navigate new situations to continue towards a desired future.

Demand for foresight is high. However, beyond focusing on producing general futures knowledge, the responsibility of foresight professionals is to increase in a practical way the ability of decision makers to use futures knowledge in current decision-making.

Changing circumstances and increasing complexity not only call for evolution and renewal of foresight practice, but also for measuring and evaluating better the results and success of foresight. Interest in foresight evaluation has indeed increased internationally in recent years. However, more understanding is needed of how evaluation can genuinely support foresight and the use of the future in postnormal times, not only from the perspectives of accountability, but especially from the perspective of usefulness for decision making, something that would lead to lessons and further development of foresight practice.

The second point: *Evaluation should always be based on the purpose and questions to which answers are sought.*

Foresight evaluation thus reverts to the question of the nature and practices of foresight: how is foresight understood and what is its purpose and aim? For whom should the evaluation knowledge be produced and how should the evaluation knowledge be used? The key question is, on what criteria will the success of foresight be assessed? *If foresight is understood as the production of knowledge about the long-term future developments for planning orientation and the preparation of action plans*, the focus of evaluation should be on the relevance, transparency and traceability of the results. *If the objective of foresight is to support the creation and promotion of transformational change in postnormal times*, evaluation should focus on analyzing the contribution of foresight to improve the capacity to create change. If the first orientation – planning and preparation - emphasizes the quality of outputs, such as foresight reports, scenarios and structural analysis, the second approach highlights more systemic understanding, openness to change and capacity to use the future for transformative change and change management. Here, capacity for future thinking and foresight literacy throughout change are more important. In both orientations, if evaluation focuses on outcome and impact, it is essential to understand how foresight as practice and knowledge producing process is linked to decisions and actions.

The third point: *Different foresight approaches require different evaluation approaches, processes and methods. It is thus necessary to be able to choose from a range of evaluation approaches and methods and to identify the right approach for the right purpose.*

It is essential to identify what kind of evaluation commissioning and implementation expertise is needed at any given time and context. As neither foresight nor evaluation are monoliths, but both include several paradigms with different ontological and epistemological thinking, it is necessary to develop a way to select the appropriate evaluation approach for any given foresight exercise. There are many different

approaches to evaluation (such as process, impact evaluation, developmental evaluation, empowerment evaluation, etc.) and they all have their own strengths and weaknesses. It is important to identify the operating environments and problems in which it is appropriate to apply which evaluation approach and methods are appropriate.

The fourth point: *We can develop a simple categorization of different types of foresight according to its focus to help frame what is relevant for evaluation*:

We have seen above two main orientations of foresight: the production of knowledge about the future for planning and strategy development and the management of transformational change in postnormal times (high uncertainty and complexity). In the first case, the focus of evaluation should be on the relevance, transparency and traceability of the results, in the second case, it should be on analyzing the contribution of foresight to the improvement of the capacity to manage that change.

In the first case, the evaluation can focus on the quality of concrete outputs, such as scenarios, lists of drivers of change, analytical parameters and the use of structured methods. In the second case, it is necessary to assess outcomes, sometimes with a large time lag between the foresight exercise and the decision-making. Also, in this second case, foresight input will be but one of the elements taken into account for decision making. This makes it more difficult to apportion the right share of impact to the foresight work that was performed. The idea of this framework is to help to identify a purposeful evaluation design and suitable methods based on the nature of foresight (see if links to the work group 4 manuals)

Framework 3: Framework for evaluating foresight

As Work Group 2 found, there is a robust literature on foresight evaluation frameworks though most have not been translated into user-friendly tools. For example, Ron Johnston developed a Foresight Impact Evaluation approach to guide foresight practitioners that can only be found in the peer review literature[10]. Below is one accessible framework for evaluating foresight adapted from a framework by Ramos et al (2019) for evaluating participatory futures. The first dimension of the framework concerns the level of impact that might be evaluated from individual to community to institution. The second dimension of the framework is based on the five stages of decision-making or purposes of foresight outlined by Ramos et al which are:

- Mapping horizons: deepening awareness of changes on medium- and long-term time horizons

[10] Johnston, R. "Developing the capacity to assess the impact of foresight." Foresight. 2012. Vol. 14, No. 1.

- Creating purpose: developing a sense of meaning and direction
- Charting pathways: creating high level strategies and socially acceptable pathways for desired change
- Acting together: mobilizing collaborative action and distributed innovation across a community to realize a desired future
- Testing ideas: generating feedback and learning about a specific idea of the future, a scenario, or prototype

		Purpose of foresight				
		Mapping horizons	**Creating purpose**	**Charting pathways**	**Acting together**	**Testing ideas**
Level of Impact	**Individual**	How did foresight deepen individuals' understanding of emerging issues within the context? How did foresight engender a greater sense of responsibility and agency for the future among individuals?	How did foresight shift awareness of the future as a tool to drive action (temporal, optimism, agency) among individuals? How did foresight make stakeholders feel more involved in decision-making?	How did foresight foster individuals' agency in strategizing about the future? How did foresight promote greater ownership among individuals over change processes (e.g. strategic planning)?	How did foresight foster individuals' agency in creating the future? How did foresight generate value and knowledge for individuals? What does this mean in this context? How did foresight help to change individual behaviors?	How did foresight support individuals to explore the impacts of this prototype on their own lives, their jobs or that of their families? How did the foresight experiment help the exploration of individual values and aspirations?
	Community	How did foresight support creative exploration of the challenges facing the community? How did foresight helps the	How did foresight create a more positive vision for the community? How did foresight create a more inclusive vision for the community? To what extent	How did foresight help the community to identify the need for change and strategies for change? How did foresight help align community values and aspirations with	How effective were the strategies for initiating change? How did foresight help the community to mobilize the energy and resources of its members	How did foresight help consider the long-term impacts of possible decisions on the community? How did foresight help make the prototype more relatable or desirable for the

Level of Impact	**Community**	community identify opportunities for change?	did it increase a shared understanding of issues, strengthen shared values, improve social cohesion?	stated priorities?	coherently for social change and sustainability?	community? How did the foresight process generate feedback from the community that led to changes or different decisions?
	Institutional	How did foresight reduce institutional blind spots? What assumptions were challenged? How did foresight help create coherence between institutional silos? How did foresight help the institution identify new opportunities?	How did foresight enhance the institution's purpose? How did foresight generate more equality / inclusion / diversity with respect to the institution's purpose?	How did foresight expand the number of options for creating change that were considered by the institution? In what ways did foresight help align institutional strategies with community needs and desires?	In what ways did foresight make the institution's actions more effective? How did foresight improve trust in the institution?	How did foresight help the institution to identify blind spots and opportunities with respect to planned action? How did participatory foresight help make adoption or rejection of an idea by the institution easier? How did foresight help the institution improve the quality of its engagement with stakeholders?

Note: The above framework is adapted from Ramos J et al (2019) Our futures by the people for the people. Available from: https://www.nesta.org.uk/report/our-futures-people-people/ Accessed: August 2022

Recommendations

We propose that the APF:

- Test the frameworks with users: The frameworks have not yet been tested with users and there has been insufficient time and resources to do this as part of this project.
- After testing, modifying, and finalizing, develop professionally designed versions of the proposed frameworks: We recognize that the presentation of the frameworks is as important as the content so professional design input is required as these skills are not within the capabilities of the working group.
- Explore the use of foresight in evaluation: The focus of this project has been the evaluation of foresight, but we see opportunities to explore the use of foresight in evaluation as this seems to be at the methodological cutting edge.

Work Group 4: Online Foresight Evaluation Toolkit

Group 4 consisted of 5 members of the Task Force:

- Eric Barela
- Annette Gardner
- Simon Schmitz
- Roger Spitz
- David Smith

Activities

The Work Group focused primarily on designing and developing an online Foresight Evaluation Toolkit, based on two online toolkit models: BetterEvaluation (https://www.betterevaluation.org/) and OECD Observatory for Public Sector Innovation (OPSI)) (https://oecd-opsi.org/guide/futures-and-foresight/). The contents were informed by Survey findings, particularly the development of two sub-sections on guidance for evaluating the foresight activities that respondents said they would like evaluation support.

The toolkit will be a curated, online collection of concepts, methods, cases, and 'how-to' resources that meet the following criteria:

- Educational: provides foresight practitioners with useful resources and guidance that support assessment, evaluative thinking, evaluation capacity building.
- User-friendly: e.g., easy to navigate, multiple door/paths structure, etc.
- Accessible, e.g., available to the public, easy to navigate, etc.
- Not exhaustive but as inclusive as possible.

- Interactive: collection and discussion of best practices that can guide practice and are a source of inspiration
- Focus on the end-user or foresight practitioner first (other users include commissioners of evaluation, evaluators new to foresight, etc.)
- Leads to something practical, e.g., insights into foresight activity–progress, quality, improved foresight practitioner excellence.

Products

A draft version of the foresight evaluation toolkit has been developed (see Appendix B) and is currently undergoing revision and will be added to the BetterEvaluation website in early 2023**.** There are two sections:

- A general 'foresight evaluation' landing page which provides definitions of foresight and evaluation constructs, as well as challenges to evaluating foresight and useful methods.
- Two sections on evaluating specific foresight activities which are practiced by many foresight practitioners as well as the ones that APF Survey respondents said they wanted evaluation support for: alternative scenarios and scanning.

Recommendations

1. APF should adopt a multi-tier model to address the diversity in APF member evaluation capacity, including 'Beginners,' 'Some experience in evaluation', and 'Expert-level in evaluation'. APF would provide a variety of targeted resources coordinated by a qualified evaluator. (see Figure 2 above).
2. APF should play a leadership role in building foresight practitioner evaluation capacity, including having a part-time dedicated staff person (or consultant) who coordinates/conducts the following tasks:
    a. Learning from foresight evaluation case studies:
        i. Accessible–people can load their own cases and provide input
        ii. Describe design, methods, process, challenges/facilitating factors and lessons learned.
    b. Provide technical assistance, e.g.,
        i. 1:1 mentoring
        ii. Webinars
        iii. Community of Practice
    c. Networking: e.g., sharing evaluation best practice
    d. Field building: Support evaluation of specific foresight approaches/products, processes, e.g., surface more evaluation cases in WFR themed issue.

APPENDICES

APPENDIX A: APF FORESIGHT EVALUATION TASK FORCE CHARTER

November 3, 2020

Purpose: This Charter describes the structure, objectives, scope of activities, and participants of the Association of Professional Futurists (APF) Evaluation Task Force. Reporting to APF, the Evaluation Task Force will support increased member evaluation capacity through sharing of evaluation resources, strategies, designs, and methods, and provide suggestions for appropriate foresight outcomes and indicators to guide evaluation design. The Task Force will operate until July 2022.

Evaluation Workgroup Responsibilities: The Evaluation Task Force aims to have a constructive discussion about evaluation strategies, designs, and methods that are appropriate for foresight broadly defined, e.g., alternative scenarios, foresight education, strategic foresight. The overarching aim is to standardize quality of foresight practice, and support achievement of foresight aims. Task Force activities include:

- Determine the challenges to evaluating foresight or 'evaluability', such as resource constraints, a long-time horizon, etc.;
- Clarify the state of evaluation practice by futurists, e.g., survey APF members to assess their work, challenges encountered, methods used, etc.;
- Identify evaluation designs, outcomes, methods, and tools that have the potential to further inform practitioner excellence in foresight; and
- Develop an 'evaluation capacity building' (ECB) model for expanding APF member evaluation expertise.

Approach: The Evaluation Workgroup is composed of a six-member group of individuals from APF. Upwards of 4 ex-officio members who are not members of APF but have expertise in foresight evaluation will also be invited to participate. This may include a representative from the World Futures Studies Federation, the American Evaluation Association, the UNESCO Foresight Department, as well as others.

The following criteria will be used to select members: 1) Members should have expertise in evaluation of foresight, including evaluation design and metrics; 2) They should be interested in identifying foresight outcomes or desired changes brought about by a particular foresight activity; 3) Finally, they should be familiar with APF foresight competencies and linking evaluation practice to increased APF member excellence.

The Evaluation Workgroup will have standing monthly calls that will be comprised of two activities: a 30-minute discussion of an issue or topic selected by the call Chair, such as the state of foresight evaluation practice, and a 30-minute technical discussion, such as relevant resources, appropriate evaluation design, etc. Outside of the monthly conference call, the Evaluation Task Force can organize ad hoc committees in order to accomplish specific objectives. The Task Force Chair will also facilitate regular communication by email, such as sharing of resources to Task Force members, trouble-shooting in between conference calls, and providing updates on Task Force activities.

Leadership and Reporting: The responsibility for chairing the calls will fall to the Evaluation Task Force Chair, Annette Gardner. She will work with the Task Force members to develop the Call Agenda and send out resources before and after the Call. She will provide a quarterly summary of its activities to APF. The Workgroup can seek approvals of policies or resources as needed. She will also provide evaluation capacity building resources to the Task Force and APF membership via the APF website.

Timeline of Activities: November 2020 to July 2022

- October – December 2020:
    - Conduct 'listening' sessions at APF activities to surface member expertise, interests, and challenges.
    - Develop Task Force member criteria for inclusion, including a combination of evaluation expertise or familiarity with evaluation design methods and foresight experience and in-depth knowledge of foresight in the areas of: alternative scenarios, education and/or training assessment, and foresight processes, such as back-casting and environmental scanning.

- January 2021:
    - Recruit and finalize 6 APF and 4 ex-officio members Task Force members

- February 2021 – June 2022
    - Monthly Task Force Calls
    - Work Group calls – TBD
    - Task Force activities, such as surveying APF membership on evaluation practices

- July 2022:
    - Produce Task Force Report and Recommendations:
        - State of foresight evaluation of APF membership—gaps and strengths
        - Appropriate Evaluation Capacity Building (ECB) resources
        - Appropriate foresight outcomes, indicators and methods to support foresight evaluation

APPENDIX B: WORK GROUP DOCUMENTS, PRODUCTS

Work Group 1: APF Foresight Evaluation Capacity Survey Instrument

Dear APF Member,

In early 2021, The Association of Professional Futurists established the APF Evaluation Task Force to increase member evaluation capacity through sharing of evaluation resources, strategies, designs, and methods. The Task Force will also provide suggestions for appropriate foresight outcomes and indicators to guide evaluation design. To achieve these aims, the Task Force is administering a short survey to the APF membership to assess the capacity of APF members to evaluate foresight activities, including strengths and gaps, as well as identify appropriate resources, such as a curated foresight evaluation bibliography. This short survey should take approximately 10 minutes to complete. If you have any questions about the survey, please contact: Bruce Tonn (btonn@threecubed.org) or Bill Lesieur (bill.lesieur@gmail.com) Please complete the survey **September 17, 2021**.

Section 1: About You and Your Organization.

1. Please indicate your APF membership category.

2. How long have you been a foresight practitioner?

- Less than 1 year
- 2-4 years
- 5-10 years
- 11-19 years
- 20+ years

3. Where are you based?

- North America
- Central America
- South America
- Africa
- Western Europe
- Eastern Europe
- Asia

4. Which description best describes your current work in the foresight field?

(Select all that apply)

- I convene foresight workshops and other types of meetings

- I conduct foresight analyses
- I procure foresight services
- I use the outputs of foresight activities
- I evaluate foresight activities
- I manage foresight activities
- I am a foresight educator
- Student
- Other *(please describe)*
- None of the above

5. Which description best describes your employment in the foresight field?

- I work for a foresight unit in a governmental organization
- I work for a foresight unit in a private sector firm
- I work for a foresight unit in a non-governmental organization
- I am a consultant who conducts foresight analyses for clients
- I convene foresight workshops and other activities for clients
- I am an academic/researcher in the field of foresight and future studies
- I am a student in the field of foresight and future studies
- Other

Section 2: You and Your Organization's Evaluation Practice

6. Do you or your organization have the ability to evaluate foresight activities?

- Yes have the ability and do evaluate foresight *(If selected, go to Q7 and Q8)*
- Yes have the ability to evaluate but do not evaluate foresight *(If selected, go to Section 3 Resources)*
- No *(If selected, go to Section 3: Resources)*
- Don't know *(If selected, go to Section 3: Resources)*

7. Who does evaluation within your organization? *(Select all that apply)*

- I do evaluation as part of a contract.
- We have an evaluator on staff.
- We have staff that includes evaluation as part of their activities.
- We work with an external evaluator.
- Other *(please describe)*

8. Why do you or your organization evaluate its foresight? *(Select all that apply)*

- Learning - to improve planning and implementation of foresight strategies and activities
- Assess outcomes and effectiveness of foresight activities
- To communicate with funders in grant reports and proposals
- To fulfill grant requirements from a funder *(or funders)*
- To change how we allocate resources
- To seek out new partners or allies for our foresight work
- To share information with policymakers and other decision-makers

- To change what polices, issues and/or organizations we focus on in our foresight work
- Other *(please describe)*

9. Who uses the results and findings from evaluations and foresight activities? *(Select all that apply)*

- Executive staff (CEO/ED)
- Managerial staff (e.g. program or project directors)
- Non-managerial staff (e.g., coordinators)
- Board of Directors
- No or limited use of results and findings from evaluation
- Funders
- Policy makers
- Partner organizations
- Other (please describe)

Evaluation activities are typically described either as process evaluations and/or impact evaluations. Process evaluations take place early in a project process and typically assess the strengths and weaknesses of foresight processes, such as the efficacy of training, or the interaction of commissions, or the workflow of distributed teams who produce those outputs. They are also referred to as 'formative evaluations.' Impact evaluations, which are also referred to as summative evaluations, assess the success of a foresight project or process and whether they made a difference. They usually include measuring the outcomes and determining the impact of a foresight initiative on decision-making, organizations, and/or communities. Please refer to these definitions when answering the next several questions

10. Which best describes your level of expertise in process evaluation?

- Mastery/Expert, e.g., I have training in process evaluation and lead foresight evaluations.
- Proficient/Skilled, e.g., I have some training and experience in doing process evaluation
- Entry/Novice, e.g., I am aware of process evaluation but no experience in doing evaluation.
- No expertise in process evaluation.
- Other – write in

11. Which best describes your level of expertise in impact evaluation?

- Mastery/Expert, e.g., I have training in process evaluation and lead foresight evaluations.
- Proficient/Skilled, e.g., I have some training and experience in doing process evaluation
- Entry/Novice, e.g., I am aware of process evaluation but no experience in doing evaluation.
- No expertise in process evaluation.
- Other – write in

12. Over the last five years, how many of the foresight activities you have been involved with have undergone process evaluations?
have undergone process evaluations? *(Check one)*

- Majority (more than 50%)
- Some (10% - 49%)
- Hardly Any (1 – 9%)
- None (Skip to Q13)

12a. Which best describes who conducted the process evaluations?

- All conducted by yourself
- Most conducted by yourself
- About half conducted by yourself and half conducted by a third party
- Most conducted by a third party
- All conducted by a third party
- Other – write in

12b. How would you rate your organization's capacity to conduct process evaluations?

- Very high
- High
- Moderate
- Low
- Non-existent

13. Over the last five years, how many of the foresight projects you have worked on have undergone impact evaluation?

- Majority (more than 50%)
- Some (10% - 49%)
- Hardly Any (1 – 9%)
- None (skip to Q14)

13a. Which best describes who conducted the impact evaluations

- All conducted by yourself
- Most conducted by yourself
- About half conducted by yourself and half conducted by a third party
- Most conducted by a third party
- All conducted by a third party
- Other – write in

13b. How would you rate your organization's capacity to conduct impact evaluations?

- Very high
- High
- Moderate

- Low
- Non-existent

Section 3: Foresight Evaluation Resources

14. What are the biggest challenges to evaluating foresight processes or products? *(Select all that apply)*

- Clients are not interested in program achievements and longer-term impacts of activities
- Insufficient financial resources
- Limited staff time
- Managing funder expectations and/or reporting requirements
- Insufficient support from organizational leadership
- Insufficient support from organizational staff
- Not knowing where / how to get started with evaluation
- Limited staff evaluation knowledge, skills, and/or tools
- Lack of appropriate outcomes, indicators, and methods that fit my organization's work
- Difficulty in developing rigorous designs, e.g., valid, generalizable
- Difficulty in specifying variables to measure impacts
- Difficulty in collecting the requisite quantitative and qualitative data
- Difficulty in analyzing data
- Difficulty in communicating results
- Lack of follow-up and use of evaluation findings, e.g., to inform strategy.
- Other *(please describe)*

15. Looking at the following list of foresight activities, which ones would you like more support with evaluation? *(Select all that apply)*

- Developing/working with alternative scenarios
- Foresight trainings and workshops
- Anticipatory futures
- Causal Layer Analysis
- Participatory futures
- Future literacy laboratories
- Environmental (or Horizon) Scanning
- Forecasting
- Quantitative analyses
- The Delphi Method
- Simulation and computer modeling
- Gaming
- Ethnographic futures research
- Other *(please describe)*

16. What types of resources can APF provide to help you evaluate your foresight activities? (Select all that apply)

- 'How-to' evaluation guide
- A repository of foresight evaluation articles, cases, and measures
- Referrals to professional evaluators
- Regular meetings of an evaluation special interest group, e.g., to share evaluation designs, measures
- Evaluation trainings, e.g., quarterly webinars, APF conference sessions
- Technical expertise from evaluators, e.g., development of an evaluation plan, theory of change, program logic models, etc.
- Other *(please describe)*

17. If applicable, please describe something you have read, seen, or heard that has helped you think about, and perhaps do, an evaluation of foresight activities? *(Please include available weblinks)*

18. If applicable, can you briefly describe one or two evaluations of foresight activities that you have undertaken?

Work Group 2 materials:

1. Zotero bibliography: https://www.zotero.org/groups/4713929/apf_evaluation_library_vs2/library
2. Tree classification of types of foresight evaluation papers
    a. As a web page https://mande.co.uk/wp-content/uploads/2021/10/tree%20diagram%20of%20papers%20Hierarchical%202022%2006%2021%20vs%201600%20wide%202.html
    b. As an Excel file

Work Group 4: Foresight Evaluation Toolkit Description

Toolkit Purpose

A curated, online collection of concepts, methods, cases, and 'how-to' resources that meets the following criteria[1] :

- Educational: provides foresight practitioners with useful resources and guidance[2] that support assessment, evaluative thinking, evaluation capacity building.
- User-friendly: e.g., easy to navigate, multiple door/paths structure, etc.
- Accessible, e.g., available to the public, easy to navigate, etc.
- Not exhaustive but as inclusive as possible.
- Interactive: collection and discussion of best practices that can guide practice and are a source of inspiration
- Focus on the end-user or foresight practitioner first (other users include commissioners of evaluation, evaluators new to foresight, etc.)
- Leads to something practical, e.g., insights into foresight activity–progress, quality, improved foresight practitioner excellence.

Target Users

- Foresight practitioners first
- Evaluators needing grounding in 'foresight' and appropriate evaluation constructs and methods.

Online Website Template—DRAFT

A. Foresight Evaluation Landing Page (modeled after BetterEvaluation Template)[11]

1. Overview of 'foresight evaluation'

a. Definition:

Foresight evaluation is the evaluation of the design, implementation, and effects of foresight activities. These activities, such as development of alternative scenarios and environmental scanning, typically entail thinking about and using the future to 'navigate those areas of life and society that are characterized by complexity and uncertainty.'[12]

b. History:

Evaluating the work of futurists is a perennial topic of concern, particularly the inability to validate foresight outputs (such as workshops and trainings, scanning processes, alternative scenario projects).[13] The nature of foresight work poses challenges,

[11] Note: the DRAFT toolkit has been reviewed by BetterEvaluation and will undergo additional revision.
[12] Shallowe, et al., 'A stitch in time?' (2020). The Royal Society for Arts, Manufactures, and Commerce.
[13] Amara, R., How to Tell Good Work from Bad. The Futurist. April 1981, 63-71.

particularly the long-term impact of foresight and desire to demonstrate attribution. That said, the divide between evaluation and its focus on demonstrating value and the foresight field's desire to demonstrate effectiveness is narrowing. Starting in the early 2000s, a growing number of futurists argued for increased evaluation to assess for quality, success, and impact while also developing foresight evaluation frameworks to guide foresight practitioners. Concurrently, the evaluation arena is paying more attention to demonstrating contribution and the role played by an initiative as one of many factors.

c. Rationale:

At one level, the futures studies discipline has wrestled with rigorously demonstrating effectiveness, undercutting its credibility as a discipline. In most cases, there is no counterfactual. Until recently, no clear guidelines and measures existed. At another level, foresight practitioners haven't made assessment of their work a high priority. However, increased sponsor/client interest in accountability and determining whether foresight outputs are useful has resulted in greater open-ness and ability to apply formal evaluation concepts and tools to foresight activities, as well as professional development activities. Increasing excellence in foresight is now a driving force for embedding evaluation in foresight. Additionally, there is synergy between foresight and evaluation, and they share constructs and methods, such as systems thinking tools, though they are applied differently in each discipline. Last, organizations and professional associations, including the Association of Professional Futurists and World Futures Studies Federation have made evaluation a priority. And, foresight organizations, such as Sitra, the Finish futures think tank, researchers, and evaluators have made significant progress in adapting evaluation constructs and designs to evaluate foresight projects and processes.

d. Purpose:

As argued by Georghiou and Keenan in one of the earliest papers on what constitutes an appropriate foresight evaluation strategy, there is no 'one-size-fits-all' approach[14]. The purpose for foresight evaluation depends on the context, which can include government, corporations, nonprofits, communities, and education. In government, the focus of foresight evaluation has been primarily on accountability and whether foresight was efficiently conducted. In foresight education, the focus is on assessing student learning outcomes and competency. Corporations and nonprofits are interested in the quality of the foresight process and/or whether it achieved its desired impacts. While differences exist, increasingly the purpose of foresight evaluation is broadening to include learning that contributes to improved field credibility, as well as improved decision-making, influencing policymaker and public awareness and understanding of the future, and greater capacity to navigate uncertainty and

[14] Georgiou, L. and Keenan, M. (2006). Evaluation of national foresight activities: Assessing rationale, process, and impact. Technological Forecasting & Social Change 73 761-777.

change. Please see APF curated foresight evaluation bibliography to search evaluation resources by foresight activity.

2. Foresight constructs useful to foresight evaluation:

A key capacity is having an understanding of the attributes of foresight that distinguish it from other topical areas or disciplines. In this section, we provide brief descriptions of core concepts that are embedded in most foresight projects and processes. Please refer to the Resources at the end for more information on foresight concepts and tools.

a. Foresight:

Foresight, the competency and practice of futures studies,[15] is the capacity to think systematically about the future to inform decision making today.[16] Specifically, it focuses on investigating the drivers of change and exploring possible futures to inform planning and policymaking. Understanding, anticipating, and navigating different types of change is a core feature of foresight work, particularly disruptive events, future developments, risks, and opportunities. Foresight practitioners use systematic methods to understand change and develop alternative scenarios, such as environmental and horizon scanning, scenario development, futures wheel, trend analysis, and forecasting. Foresight does not seek to predict since none of the futures that come to pass are exactly as imagined. However, it does support better preparation for any future which may arise, as well as spur imagination and collective creativity.

b. Futures Literacy and anticipation:

Futures literacy is a capability that is acquired by individuals and is fundamentally social, collective, and relational in its dynamics. It allows people to better understand the role that the future plays in what they see and do. The Futures Literacy Laboratory (FLL) is a method for helping people learn how to use the future through collective knowledge creation, enabling them to anticipate future problems and conditions.[17]

c. Futures plural:

Developing and/or working with alternatives scenarios or stories about the future is a staple of many foresight practitioners. Rather than assume a specific event, outcome, or future, foresight prepares for many eventualities and multiple possible futures that are well beyond the usual time horizon of a program evaluation. Seeing the world through multiple lenses helps to make choices that take into consideration changing conditions as well as manage complexity and ambiguity. While near-term aspects of

[15] Shallow, A. et al. A stitch in time? Realizing the value of futures and foresight. RSA. October 2020.

[16] Maree Conway, "**Foresight: an introduction**", page 2,

[17] Miller, R. What is Futures Literacy? What is a Futures Literacy Laboratory? A brief introduction. Presentation. 2022.

a scenario process lend themselves to evaluation (such as changes in participant understanding), most scenarios have a time horizon well beyond that of a typical program evaluation.

d. Foresight professional competencies:
The field is adopting professional competency models, such as the Association of Professional Futurists' six-practices Foresight Competency model, which include:[18]

I. Framing: scoping the project, defining the focal issue and current conditions
II. Scanning: exploring signals of change or indicators of the futures
III. Futuring: identifying a baseline and alternative futures
IV. Visioning: developing and committing to a preferred future
V. Designing: developing prototypes, offerings or artifacts to achieve the vision and goals;
VI. Adapting: Enabling organizations to generate options to alternative futures.

These competency models point to areas of assessment of professional expertise, in addition to examining program or project success.

3. Tackling Foresight Evaluation Challenges

a. Foresight initiative complexity:
Foresight approaches such as alternative scenarios are designed to characterize and suggest action steps in the face of complexity. They are also themselves complex initiatives, engaging many stakeholders with diverse perspectives to learn and use a new way of thinking. An important feature of foresight is to understand the features of complexity and get comfortable with nonlinearity. Complex problems have many components that interact in ways that may be inexplicable. Complex adaptive systems models are helpful for understanding and anticipating complex problems and situations. Also, the evaluation arena has been addressing complexity for some time and resources can be found at: https://www.betterevaluation.org/en/themes/complexity

b. Emergent nature of foresight initiatives:
Since the future hasn't happened yet, and much of foresight activities work is by definition 'emergent.' Emergence refers to the process whereby the existence or formation of novel collective behaviors, properties, or phenomena come into existence only when the parts of a system interact in a wider whole. A term often used in philosophy, systems thinking, and sciences, emergence generates synergies between the individual parts, which would not arise unless they interact. It's important to understand the role of emergence in foresight and consider approaches like developmental evaluation, as well as identify and document emergent partners,

[18] APF. Foresight Competency Model. August 2016.

strategies and outcomes, rather than only paying attention to the objectives and targets identified at the beginning.[19] Likewise, it is important to identify the aspects of a foresight initiative that are not emergent and can be approached using conventional evaluation methods, such as assessing participant satisfaction, changes in knowledge, and fidelity of implementation.

c. Contribution vs. Attribution:
Because the causal chain between a foresight intervention and impact can be unclear and long, tools like systems maps, contribution analysis, and process tracing that look at contribution are more useful. Similar to other complex and uncertain evaluands, such as advocacy initiatives that also have a long time-horizon and multiple stakeholders, foresight evaluators need to educate stakeholders on the challenges of attribution and the value of focusing on contribution.

d. Rigor (generalizability, validity):
While determining attribution is a challenge, foresight evaluations can be designed to increase rigor, including collecting baseline data for a pre/post or longitudinal analysis, triangulation of data from diverse sources—interviews, sites visits, surveys, financial analyses--to corroborate (or refute) findings, having a reasonable sample size and response rate in surveys, identifying a counterfactual (such as survey with placebo technique), and inquiring about unintended consequences.

e. Managing stakeholder expectations:
Consumers of foresight processes and products may have unrealistic expectations on the rigor and certainty of foresight evaluation findings. Also, given the long-time horizon of foresight outputs, such as forecasts and alternative scenarios, stakeholders may question the relevance of evaluation findings that focus on short-term outcomes in decision-making.

4. Useful evaluation concepts and methods:

In this section, we highlight those evaluation constructs and methods that are well suited for evaluating foresight, addressing the challenges noted above and providing evaluation findings that focus on process and impact. (They are listed alphabetically, not in order of importance.

Concepts:

a. Adaptive management:
The term 'adaptive management' refers to adaptation that goes beyond the usual adaptation involved in good management - modifying plans in response to changes in circumstances or understanding and using information to inform these decisions. Adaptive management refers to an approach to managing under conditions of

[19] BetterEvaluation web-site.

ongoing uncertainty which represents a paradigm shift from classic, linear approaches to planning, implementation and evaluation. Source: BetterEvaluation.

b. Appreciative inquiry:

Appreciative Inquiry is an approach to organizational change which focuses on strengths rather than on weaknesses. Appreciative Inquiry is often presented in terms of a 4-step process around an affirmative topic choice:

1: DISCOVER: What gives life? What is the best? Appreciating and identifying processes that work well.

2. DREAM: What might be? What is the world calling for? Envisioning results, and how things might work well in the future.

3. DESIGN: What should be--the ideal? Co-constructing - planning and prioritizing processes that would work well.

4. DESTINY (or DELIVER): How to empower, learn and adjust/improvise? Source: BetterEvaluation.

c. Complex adaptive systems (CAS):

A complex adaptive system is a system that is *complex* in that it is a dynamic network of interactions, but the behavior of the ensemble may not be predictable according to the behavior of the components. It is *adaptive* in that the individual and collective behavior mutate and self-organize corresponding to the change-initiating micro-event or collection of events.[1][2][3]

d. Developmental evaluation:

Developmental Evaluation (DE) is an evaluation approach that can assist social innovators develop social change initiatives in complex or uncertain environments. DE originators liken their approach to the role of research & development in the private sector product development process because it facilitates real-time, or close to real-time, feedback to program staff thus facilitating a continuous development loop. Source: BetterEvaluation.

e. Empowerment evaluation:

Empowerment evaluation is a stakeholder involvement approach designed to provide groups with the tools and knowledge they need to monitor and evaluate their own performance and accomplish their goals. It is also used to help groups accomplish their goals. Empowerment evaluation focuses on fostering self-determination and sustainability. It is particularly suited to the evaluation of comprehensive community–based initiatives or place-based initiatives. Source: BetterEvaluation.

f. Evaluative thinking:

A mind-set that goes beyond perfunctory assessment of a program's value and involves: ' identifying assumptions, posting thoughtful questions, pursuing deeper

understanding through reflection and perspective taking, and making informed decisions and preparation for action.'[20]

g. Systems thinking:
"Systems thinking, in the evaluation field, often refers to a way of thinking based on core systems concepts. To date, three distinct orientations to systems thinking: One orientation draws from historical review and identifies interrelationships, perspectives, and boundaries as core concepts of focus present in much of system theory (Williams and Imam, 2007); A second orientation draws from the field of cognitive science to identify processes for thinking that focus on distinctions, relationships, perspectives, and boundaries (Cabrera and Cabrera, 2015); and The third orientation draws on human systems dynamics theory and focuses on concepts of containers, differences, and exchanges (Eoyang, 2007)."[21]

h. Transformational evaluation:
Paper on the BetterEvaluation website: "'evaluation must consider all interventions in their broader context and how they interact with human and natural systems', given their significant impacts. Evaluation also needs to move beyond a focus on individual projects and their stated objectives to consider their impact on wider systems. Source: BetterEvaluation.

Methods:

a. Contribution Analysis:
Contribution Analysis is an approach for assessing causal questions and inferring causality in real-life program evaluations. It offers a step-by-step approach designed to help managers, researchers, and policymakers arrive at conclusions about the contribution their program has made (or is currently making) to particular outcomes. It is particularly useful in situations where the programme is not experimental. Source: BetterEvaluation.

b. Evaluability Assessment (EA):
An assessment of the extent to which an intervention can be evaluated in a reliable and credible fashion. The results of an EA should indicate what preparatory steps are needed to make an intervention ready for an evaluation and what remaining challenges will need to be addressed by an evaluation ream. An EA will examine: (a) the adequacy of the intervention's Theory of Change, (b) the availability of relevant data and supporting systems, (c) stakeholders' interests in the evaluation, especially

[20] Vo, AT, Schreiber, JS, and Martin, A. Toward a Conceptual Understanding of Evaluative Thinking. New Directions For Evaluation. No. 158. Summer 2018. Pg. 31. Citing Buckley et al. 2015.

[21] Principles for Effective Use of Systems Thinking in Evaluation Systems in Evaluation TIG A Topical Interest Group of the American Evaluation Association Revised September 9, 2018

their evaluation questions, (d) constraints arising from the surrounding institutional context. Source: BetterEvaluation.

c. Monitoring:
Monitoring is a process to periodically collect, analyze and use information to actively manage performance, maximize positive impacts and minimize the risk of adverse impacts. It is an important part of effective management because it can provide early and ongoing information to help shape implementation in advance of evaluations. Source: BetterEvaluation.

d. The Most Significant Change technique:
A form of participatory impact monitoring and participatory evaluation, which involves soliciting and analyzing personal accounts of change and deciding which of these accounts is the most significant – and why. Learning opportunities occur at two levels: (a) the choices of which types of change are more versus less preferred i.e. the direction of change, (b) the choice of what criteria of value are most important in a given context i.e. the nature of the objectives being pursued by those changes. In addition, the way in which MSC story selection processes are structured enables differences and similarities in views of different stakeholder groups to be more visible (relating to both levels of learning) Source: MandE NEWS

e. Success Case Method:
The Success Case Method (SCM) involves identifying the most and least successful cases in a program and examining them in detail. It is a useful approach to document stories of impact and to develop an understanding of the factors that enhance or impede impact. The Success Case Method deliberately looks at the most, and least, successful participants of a program. The purpose is not to examine the average performance - rather, by identifying and examining the extreme cases, it asks: 'When the program works, how well does it work? What is working, and what is not? Source: BetterEvaluation.

5. **Foresight evaluation resources, e.g.,**
   a. **The Futures Toolkit: Tools for Futures Thinking and Foresight Across UK Government**. Government Office for Science. Edition 1.0. November 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674209/futures-toolkit-edition-1.pdf
   b. Shallowe, A., Szymczk, A., Firebrace, E., Burbridge, I. and Morrison, J. **A stitch in time? Realizing the value of futures and foresight.** Royal Society for Arts, Manufactures and Commerce (RSA). October 2020. https://www.thersa.org/globalassets/_foundation/new-site-blocks-and-images/reports/2020/10/rsa-stitch-in-time.pdf
   c. APF Curated Foresight Evaluation Bibliography
      i. On Zotero online
      ii. On LitMaps'

d. Appropriate evaluation 'how-to' guides, e.g.,
   i. Robinson, S. Professional Development Program Evaluation for the Win. https://www.frontlineeducation.com/program-evaluation/
e. Discussion papers, e.g., state of foresight evaluation:
g. Online: instructional webinars, guides
h. Foresight trainings programs: There are diverse training options, including:
   i. **Youth: Teach the Future**:
   A nonprofit organization, founded by leading futurist Peter Bishop, dedicated to bringing foresight and futures thinking to learners around the world. Teach the Future provides practical workbooks, resources, and strategies prepared to help learners of all ages, anywhere.
   ii. **Broader Audiences: UNESCO Futures Literacy Program:**
   The United Nations agency responsible for education - has driven futures literacy as a key skill for the 21st century. It conducts Futures Literacy Labs (FLL) around the world.
   iii. **Institute for the Future Futures Thinking Online course**:
   The online self-study course offered by the platform Coursera offers a comprehensive introduction to the benefits and approaches of Futures Thinking. In a specialization of five courses, participants can learn how to use the most important methods and adapt their mindset to Futures Thinking.
   iv. **DLR Project Management Agency (DLR-PT) Courses on Strategic Foresight**:
   Directed towards organizations or groups of stakeholders that seek to strengthen their knowledge and skills on Strategic Foresight, the DLR-PT is offering capacity-building workshops. In these tailored courses, participants can experiment with the most important methods of Strategic Foresight and learn how to integrate them in their daily tasks.
   v. Futures & Foresight Studies: Academic Programs & Courses
   Universities:
   Arizona State University, The Center for the Study of Futures (USA); Ontario College of Art & Design University, Strategic Foresight & Innovation MDes (OCAD U Canada); Tamkang University, Graduate Institute of Futures Studies (Taiwan); University of Hawaii at Manoa (USA); University of Houston, M.S. in Foresight (USA); Oral Roberts University, Doctor of Strategic Leadership in Foresight (USA); University of Notre Dame, Foresight in Business & Society, Mendoza College of Business (USA); University of Stellenbosch Business School (USB) - Institute of Futures Research (IFR, South Africa); University of Turku, Turku School of Economics: Finland Futures Research Centre (Finland)

Other Courses: Metafuture Online Futures School (Become a Futurist with Sohail Inayatullah, Australia), The Futures School (pioneered by Kedge, USA)

WEBSITE SUB-SECTIONS: SPECIFIC FORESIGHT ACTIVITIES

The following two sections provide guidance on evaluating the foresight activities that are most common and were rated as those that foresight practitioners wanted evaluation support.

A. Evaluating Alternative Scenarios

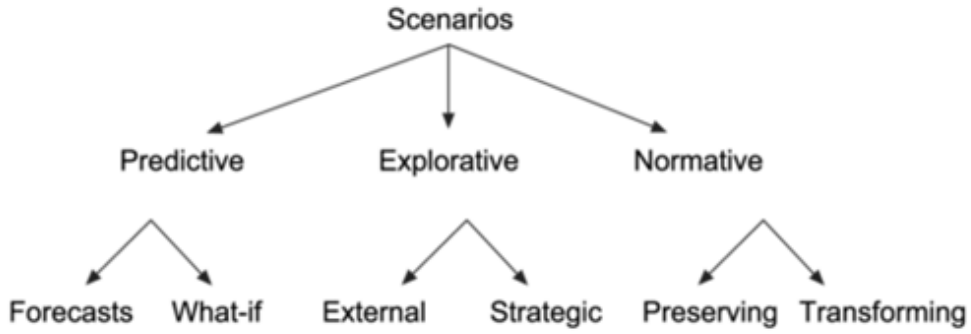
Definition of (alternative) scenarios

According to the UNDP's Global Center for Public Service Excellence, scenarios are "stories (or narratives) set in the future that explore how the world would change if certain trends were to strengthen or diminish, or various events were to occur. Scenario planning does not attempt to predict what will happen, but through a formal process identifies a limited set of examples of possible futures that provide a valuable point of reference when evaluating current strategies or formulating new ones." (UNDP, 2015).

Which different types of alternative scenarios exist?

One frequently used categorization of scenario techniques has been introduced by Börjeson *et al.* (2006). The authors differentiate between **predictive scenarios** (What will happen? -> forecasts & what-if scenarios), **explorative scenarios** (What can happen? -> external and strategic scenarios), and **normative scenarios** (How can a specific goal be reached? -> preserving or transformative such as back-casting).

Figure 1. Categorization of scenarios, according to Börjeson et al. (2006)



What are techniques to elaborate alternative scenarios?

Bishop et al. (2007) developed eight categories/types of scenario techniques (with two to three variations for each type). They are presented only briefly here as readers are encouraged to consult Bishop et al. 's article for a detailed instruction on how/when to use the respective scenario technique.

1. **Judgment techniques** rely mostly on an individual or a group's judgment on how the future could unfold ("thinking the unthinkable", using visualization, role playing or judgmental forecast)

2. **Baseline futures** elaborate one scenario which describes the expected future ("The most likely future isn't"). Variations include the Manoa technique by Wendy Schultz (1993) and the Systemic Scenarios (Burchsted and Crews, 2003).

3. **Fixed scenarios** are decided ahead of time, whereby participants are asked to elaborate the simplest statement of the scenario's essence. Variations include Incasting (Schultz, n.b.) and the Stanford Research Matrix (Hawken et al., 1982)

4. **Event sequences** use future series of events ("future branches"). Alike the past, we can think of the future of series of events with the exception that future series are yet unknown. If a potential development occurs, the future branch turns one way, if not, then another way. Variations of these probability trees can be identified: "one uses the branches to create scenario themes and the other builds the sequences after developing the events" (Lisewski, 2002; Buckley and Dudley, 1999; Covaliu and Oliver, 1995).

5. **Backcasting**, as opposed to forecasting, envisions a future state at a specific time horizon, from which one traces back the respective potential developments to the present (Robinson, 1990). Variations: Horizon mission methodology, future mapping, impact of future technologies

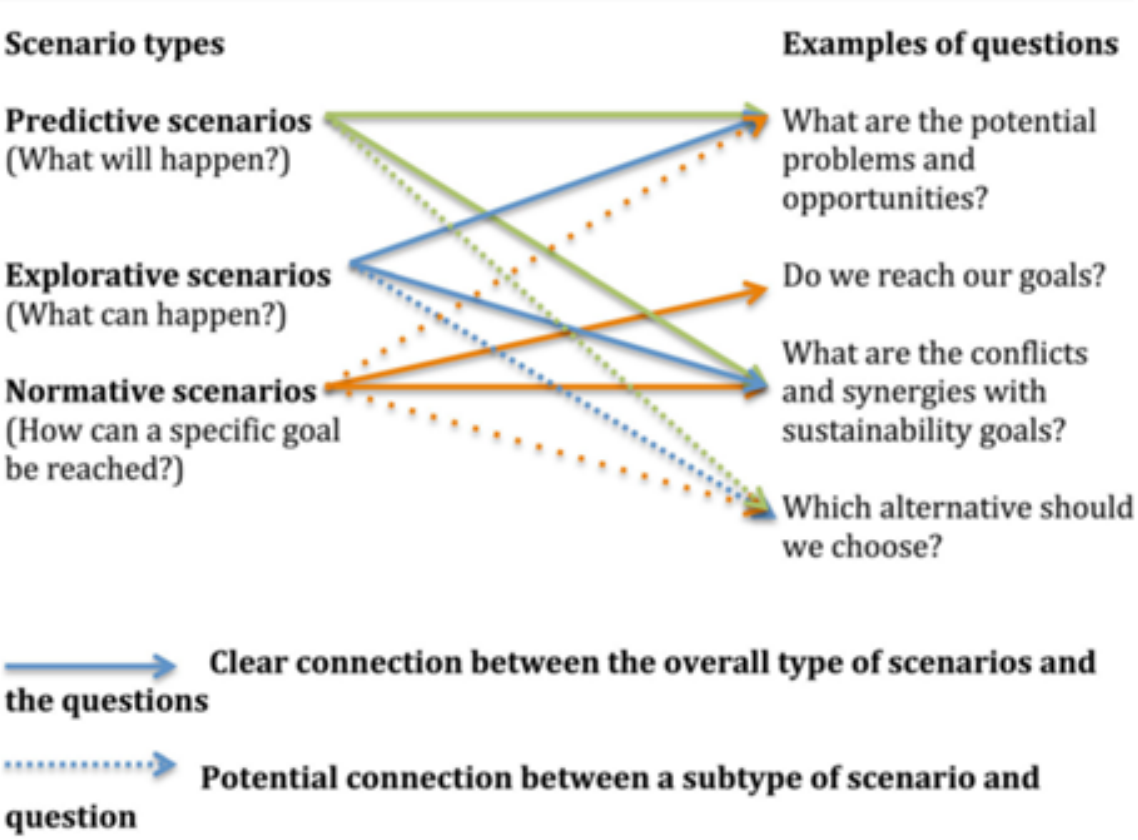
6. **Dimensions of uncertainty**: In this approach, sources of uncertainties are identified, which subsequently are used as a basis for the development of scenarios according to the uncertainties' developments. Variations include the Royal Dutch Shell/Global Business Network (GBN) Matrix (as the most commonly used scenario technique), Morphological analysis (MA), field anomaly relaxation (FAR), Option Development and Option Evaluation (OS/OE), or the computer programme MORPHOL

7. **Cross-impact analysis** uses a matrix format and consistently estimates the probabilities of various events and their respective influence on each other.

How can alternative scenarios be evaluated?

The evaluation methods used to assess scenarios differ according to the type of scenario. While quantitative tools that use databases may be more useful in order to examine scenarios within a shorter time horizon, qualitative assessment methods might be more applicable to assess long-term transformative scenarios (Fauré et al., 2017). Drawing on Börjeson et al.'s categorization of scenarios, Fauré et al. (2017) introduced sample evaluation questions and approaches (Figure 2).

Figure 2. Example of questions according to the scenario type



Source: Fauré et al. (2017)

How can we evaluate predictive scenarios?

Predicative scenarios try to find answers to the question "What will happen?", using forecast methods and "What if scenarios". They tend to be short-term types of scenarios, which could be evaluated using **quantitative tools such as databases**. In these types of scenarios, aiming for accuracy and undertaking an uncertainty analysis are crucial aspects (Fauré et al, 2017). In order to evaluate more "what if" scenarios, in which the different strategies and actions are taken by the respective scenario user, **Fauré et al. suggest a prioritization of alternatives and the use of tools to aggregate impacts, such as Multi Criteria Assessment** [1].

How can we evaluate explorative scenarios?

Explorative scenarios focus on the contextual environment, thus focusing on the external factors that are beyond the control of the intended users (van der Heijden, 1996). They provide inputs to examine strategic scenarios and can fulfill functions similar to an uncertainty analysis. Such a scenario tends to have a medium- to long-term horizon and depict profound changes vis-à-vis the current situation. Explorative scenarios are useful to image the implications of alternative strategic decisions and resulting actions (Fauré, 2017).

Evaluating explorative scenarios to determine X and Y could be done by using a variety of **questionnaires** and the **combination of different tools**. Arushanyan et al. (2017) suggest the use of qualitative assessments of environmental aspects within the Sustainability Assessment Framework Tool. Accordingly, the **Sustainability Assessment Framework for Scenarios (SAFS)** is a methodological framework which depicts the approach to qualitatively assess future scenarios (Fauré et al., 2017). SAFS are conducted using the following steps: scoping, inventory analysis, assessment of risks and opportunities, and interpretation (ibid.). The focus on qualitative methods by SAFS makes it a good fit to evaluate long-term scenarios that imply large transformative changes (Arushanyan et al., 2017). A different way to evaluate scenarios would be the **development of new datasets tailored to the possible external developments** (using e.g. LCA and IOA databases). One limitation with evaluating explorative scenarios is however the level of accuracy due to the large degree of change as well as comparably great uncertainty (Fauré et al., 2017).

How can we evaluate normative scenarios?

The question "How can a specific goal be reached?" is at the core of normative scenarios (Börjeson et al., 2006). These can be further differentiated in two types: "profound changes" are required for transformative scenarios while "adjustments to the current situation" are described in preserving scenarios (ibid). Such transformative scenarios usually have a long-term time horizon with a significant alteration vis-à-vis the status quo, which makes a quantitative evaluation challenging. Fauré et al. (2017) suggest using a **spread-sheet model**, which was used by Francart et al. (2016) to evaluate four scenarios for the built environment and meeting greenhouse gas emission targets in Sweden by 2050.

Additional resources to evaluate alternative scenarios

Curated bibliography compiled by members of the Association of Professional Futurists (APF)

The Task Force on Foresight Evaluation of the Association of Professional Futurists has gathered a rich collection of readings on the evaluation of foresight initiatives- including on alternative scenarios. The curated bibliography provides a more detailed overview of types of scenarios, different techniques and purposes on how scenarios can be used and a variety of approaches to evaluate them.

References used in this section

- Arushanyan Y, Ekener E, Moberg Å (2017) Sustainability assessment framework for scenarios – SAFS. Environ Impact Assess Rev 63:23–34. https://doi.org/10.1016/j.eiar.2016.11.001
- Baard et al (2012), Scenarios and sustainability: tools for alleviating the gap between municipal means and responsibilities in adaptation planning. Local Environ 17:641–662. https://doi.org/10.1080/13549839.2011.646969

- Bishop, P., Hines, A. and Collins, T. (2007), "The current state of scenario development: an overview of techniques", Foresight, Vol. 9 No. 1, pp. 5-25. https://doi.org/10.1108/14636680710727516
- Börjeson L, Höjer M, Dreborg KH, Ekvall T, Finnveden G (2006) Scenario types and techniques: towards a user's guide. Futures 38: 723–739. https://doi.org/10.1016/j.futures.2005.12.002
- Buckley, J. and Dudley, T. (1999), ''How Gerber used a decision tree in strategic decision-making'', The Graziadio Business Report, available at: http://gbr.peperdine.edu/993/tree.html.
- Burchsted, S. and Crews, C. (2003), ''Systemic scenarios: creating synergy through scenarios and systems thinking'', unpublished Master's thesis, University of Houston-Clear Lake, Clear Lake, TX
- Covaliu, Z. and Oliver, R. (1995), ''Representation and solution of decision problems using sequential decision diagrams'', Management Science, Vol. 41 No. 12, pp. 1860-81.
- Fauré E., Arushanyan Y., Ekener E., Miliutenko S., Finnveden G. (2017), Methods for assessing future scenarios from a sustainability perspective, European Journal of Futures Research (2017) 5: 17 https://doi.org/10.1007/s40309-017-0121-9
- Francart N., Malmqvist T., Hagbert P. (2016), Climate implications of a collaborative economy scenario for transportation and the built environment. KTH Royal Institute of Technology, Stockholm, https://www.researchgate.net/publication/321761687_Climate_target_fulfilment_in_scenarios_for_a_sustainable_Swedish_built_environment_beyond_growth
- Hawken, P., Schwartz, P. and Ogilvy, J. (1982), Seven Tomorrows: Toward a Voluntary History, Bantam Books, New York, NY.
- Lisewski, M. (2002), Project Proposal Influence Diagrams in Think Tools, Think Tools, Munich.
- Robinson, J. (1990), ''Futures under glass: a recipe for people who hate to predict'', Futures, Vol. 22 No. 8, pp. 820-3.
- Schultz, W. (1993), ''Scenario building: the Manoa approach'', available at: www.infinitefutures.com/tools/sbmanoa.shtml (accessed April 2022).
- Schultz, W. (n.d.b), ''Scenario incasting: exploring possible alternative futures'', available at: www.infinitefutures.com/tools/inclassic.shtml (accessed April 2022).
- UNDP (2015), Foresight Manuals- Empowered Futures, https://www.undp.org/publications/foresight-manual-empowered-futures
- Van Der Heijden K (1996) Scenarios: the art of strategic conversation. Wiley, Chichester

B. Evaluating Scanning

Definition of scanning: "…a technique for detecting early signs of potential important developments through systematic examination of potential threats and opportunities…" (OECD). It is an organized approach to looking for early signals of change—warning signs—coming in the future. It is similar to looking through a pair of binoculars to see what is

coming in the distance but with more intentionality in what is being looked for and the time horizon being used. There are two kinds of scanning though both have similar aims or assessing whether an organization or government is adequately prepared for the future and able to take effective action:

- Environmental scanning or scanning an organization's external, operating environment. It is more commonly used in North America.
- Horizon scanning is more commonly used in Europe and in some situations may mean the same thing as 'Environmental scanning.' In some cases, 'horizon' may refer to looking beyond an organization's operating context.

Types of emerging change that scanning systems look for include:

- Trends that are in the early stages of their development and discontinuities in established trends;
- Weak signals or emerging issues of change that develop into a trend
- Developments and events, including wildcards or low-probability, high-impact events and black swans or unknowable, random events; and
- Innovations in the early stages of their development.

These changes can be classified as opportunities or threats, as well as categorized by trend frameworks such as STEEP (social, technology, environment, economy, and policy).

History of scanning:

French futurists are credited with developing the concept of 'future-bearing facts' that became a founding concept in futures studies and the basis of the 'weak signal,' an American concept. Scanning also has its roots in the business environment as established by Aguilar's 1967 book, *Scanning the Business Environment*. Over the years, scanning has evolved to include different types of scanning (such as formal and informal), detailed models of scanning, targeted scanning (such as Issues Management), and broadening of scope to include wild-cards. Chun Wei Choo's article, "The Art of Scanning the Environment" (2005) provided the definitions and parameters for environmental scanning, including four modes of scanning: Undirected viewing, Conditional viewing, Informal search, and Formal search, as well as research on the effectiveness of scanning in supporting organizational learning in business arena. Since then, there has been much progress in developing weak signal classification systems, searching for wild-cards and black swans, and clarifying the definition and discernment of 'weak signals.'

Purpose of scanning:

The aim is to identify early signs of change that could have a significant impact, support preparedness and increased resilience (Bishop and Hines, 2012). It deepens the

understanding of driving forces, identifies gaps in knowledge or new areas of research, and supports a dialogue among stakeholders on key issues and solutions.

Types of Scanning Processes:
Scanning approaches have been categorized in different ways. They can be of four kinds: Undirected viewing or 'touring;' Conditional viewing or 'tracking;' Informal search or 'satisficing;' and Formal search or 'retrieving.' They may be focused or broad, or both, depending on the information needs and the purpose of the scan. They may be continuous or periodic though static scans can become outdated quickly.

While scanning processes may range in the number of steps, they typically include three stages or: 1) Signal detection and identification of 'scanning hits' through a systematic review and collection of a range of sources, depending on the purpose of the scan. The challenge is to find the signal amongst the noise or filtration; 2) Scanning items are organized and analyzed for credibility, novelty, impact, likelihood, and relevance, such as scoring and ranking them. Ideally, the scanning system is a participatory, group activity and includes experts with content expertise as well as generalists. Group processes, such as workshops, may be used to assess the hits and/or develop deeper insights than might not otherwise be surface; 3) Last, the intelligence gathered from the scanning system is used to inform decision-making, such as the impact of developments on particular policies and create new, resilient strategies. It may also be used to enable an alternative scenario process. Increasingly, an electronic platform is used to automate the data gathering and storage stage, as well as analyze the data for credibility.

Futures studies theories important for developing a theory of change.
Graham Molitor developed the emerging issues life cycle curve to describe the trajectory of an issue as it goes from being below the surface of general public awareness to being framed and making it on the public agenda and the mainstream. The model has evolved to mirror the S-curve and adoption of innovation, as well as identify the sources of information at each stage of the model. Scanners want to look as far upstream on the curve as possible, including being on the lookout for wild-cards and reviewing 'fringe' sources.[22]

The Three Horizon Model provides short-, medium-, and long-term time scales of outlook against which the development of 'weak signals' can be played out into the future. Horizon 1 are impacts that will be felt today. Horizon 2 speaks to trends whose impact will be felt in the short- to medium term. Horizon 3 is for those trends that will have an impact in the longer-term.

[22] Molitor, G. T. T. (2018, March). The Molitor model of change. World Futures Review, 10(1), 13-17.

Key evaluation concepts and questions

Evaluation designs can be adapted to the three stages of the scanning process or: 1) gathering information to identify 'weak signals,' 2) Sense-making and analyzing scanning items for impact, relevance, etc.; and 3) Using this intelligence to inform decision-making on specific policies and possibly other foresight activities. Designs may focus solely on process and the standing up and monitoring of a tracking system or developmental and/or process evaluation. Alternatively, designs might focus on process and outcomes/impacts, depending on stakeholder information needs. Conventional evaluation designs and methods apply in most cases.

Stage 1: evaluation activities focus on the implementation and quality of the scanning process or formative evaluation. Evaluation questions include:

- Implementation: did the tracking system get set up as intended? If not, why not
- What were the challenges (such as cost, administrative burden) and facilitating factors (such as leadership buy-in, automation) in setting up the process?
- Developmental: how has the tracking system evolved? What refinements were made?
- Quality:
    - How well did the system identify scanning items?
    - How well did the system monitor scanning items?
    - How can the process be improved?

Stage 2: evaluation activities focus on the analysis of scanning items and collective learning or 'sense-making' – summative evaluation. Evaluation questions include:

- Quality of the scans—are they complete or are there gaps? What is the validity of scans?
- Relevance of the scans—do they identify important risks and/or opportunities?
- How well did the system assess the potential for high-impact scanning items?
- Quality of the analysis process: how can it be improved?
- Quality of the collective process to develop insights, findings—what was the level of engagement? Were diverse perspectives represented?

Stage 3: evaluation activities focus on the impact of scanning knowledge/outputs on decision-making or summative evaluation:

- How were scanning reports used, e.g.,
    - Inform funding decision-making
    - Identify or prioritize policy issues
    - Support futures discussions at different levels
    - Stay up to date on topics of interest to organization, community
    - Other

- Did scanning information increase decision-maker imagination and futures literacy?
- How likely are decision-makers to use scanning reports in the future?
- Did scanning information result in a change in X (policy, decision, funding)?
- Did scanning increase organization/government resilience and preparedness?

Post-scanning: Evaluation activities focus on the value of scanning:

- The accuracy of scanning projections[23]—did they happen and within the identified timeframe?
- What prioritization approach produced the most accurate results?

Tackling Scanning Evaluation Challenges

A perennial question is the quality of the scanning 'hit' and identifying the signal from the noise. There are a couple of approaches to discern whether an important 'hit' was overlooked, including a document analysis of reports and studies on late-identified items and a determination of when an item should have been identified. Second, experts can be interviewed on the accuracy, completeness, and usefulness of an item.

Determining the impact of a scanning process on decision-making is difficult since a scanning report is one of many inputs that a policymaker takes into consideration. Demonstrating attribution would be very difficult. One approach is to interview policy-maker staff who are frequently tasked with reviewing reports and have content expertise. Also, focusing on contribution could be a fruitful area of inquiry, demonstrating the relative value of scanning information to other sources and types of information.

Last, assessing the accuracy of prediction is possible in some areas, such as emerging healthcare technology, which is a well-documented arena, but findings are difficult to validate.

Learning from scanning evaluation best practices:

While there are many private and public sector scanning programs, with the earliest dating back to the 1970s, there are few program evaluations in the literature that can support evaluation planning. Habergger's article reviewing the public sector scanning processes in the UK, Singapore, and the Netherlands provides key success factors, such as interoperable working environments to support cross-pollination of new ideas. Hines and Bengston describe the design and early-stage development and implementation of a horizon scanning system established for the U.S. Forest Service. There are discrete studies that focus on the rigor of and trustworthiness of the scan results, and whether a scanning process captures useful futures items. And there are macro-level evaluations. Hines et al. assess scanning in healthcare, including whether scanning reports are used in policymaking.

[23] Note: this is being done in the health care technology arena. It may not apply to all arenas.

One helpful guide is the Evaluation of the AHRQ Healthcare Horizon Scanning System conducted by Mathematica Policy Research for ECRI Institute, which focused primarily on the collection and analysis of emerging health care innovations. The Evaluators used a mixed-methods design: semi-structured interviews with staff and domain experts, review of Status Update Reports, Survey of experts to assess quality and usability of intervention reports, and a survey of stakeholders on credibility and usability of Intervention reports.

Resources

- Bengston, D.N. Horizon Scanning for Environmental Foresight: A Review of Issues and Approaches. United States Department of Agriculture. September 2013.
- Bishop, PC. And Hinds, A. Teaching about the Future. Palgrave Macmillan: London, UK. 2012.
- Choo, Chun Wei., The Art of Scanning the Environment. Bulletin of the American Society for Information. 2005. 25(3), 21-24. https://doi.org/10.1002/bult.117
- Cranfield University. Action B4: Horizon Scanning Toolkit. Smarter Regulation of Waste in Europe LIFE SMART Waste Project. January 31, 2018.
- Government Office for Science. Tools for Futures Thinking and Foresight Across the UK Government. Edition 1.0. November 2017.
- Habegger, B. Strategic foresight in public policy: Reviewing the experiences of the UK, Singapore, and the Netherlands. Futures 42 (2021): 49-58.
- Hines, A., Bengston, DN, Dockry, MJ. And Cowart, A. Setting Up a Horizon Scanning System: A U.S. Federal Agency Example. World Futures Review. 2018. Vol. 10(2) 136-151.
- Hines, P., Yu, LH, Guy, R., Brand, A. and Papaluca-Amati, M. Scanning the horizon: a systematic literature review of methodologies. BMJ Open. 2019;9”e026764. Doi:10.1136/bmjopen-2018-026764.
- Institute of Risk Management. Horizon Scanning: A Practitioners Guide. Innovation Special Interest Group of the Institute of Risk Management. Year.
- National Academies of Sciences, Engineering, and Medicine 2020. Safeguarding the Bioeconomy. Washington, DC: The National Academies Press. https://doi.org/10.17226/25525.
- Rossel, P. Early detection, warnings, weak signals and seeds of change: A turbulent domain of futures studies. Futures 44 (2021): 229-239.
- Schultz, W. The cultural contradictions of managing change: Using horizon scanning in an evidence-based policy context. Foresight. Vol. 8. No. 4. 2006.