CGIAR | Independent Advisory and Evaluation Service

Mapping Evaluation Management Practices in International Research and Development Organizations

Irene Toma, Ibtissem Jouini and Daniela Maciel Pinto

September 2025

Citation: Toma, I., Jouini, I., and Maciel Pinto, D. 2025. Mapping Evaluation Management Practices in International Research and Development Organizations. Rome: IAES Evaluation Function.

Cover image: IAES Stock Image

Mapping Evaluation Management Practices in International Research and Development Organizations

Irene Toma, Ibtissem Jouini and Daniela Maciel Pinto

September 2025

Acknowledgments

The study was led by Ibtissem Jouini, Senior Evaluation Manager of the IAES Evaluation Function. Independent Consultant Irene Toma was responsible for survey administration, results analysis, and mapping of independent evaluation functions. Daniela Maciel Pinto, Analyst at the Brazilian Agricultural Research Corporation (Embrapa), conducted the literature review and contributed analytical insights to the study. Jouini also prepared the executive summary, introduction, context, and conclusions.

Overall oversight was provided by IAES Director Allison Grove Smith and Evaluation Function Lead Svetlana Negroustoueva. Colleagues and partners who participated in piloting the survey include Ann Marie Castleman, Anne Clemence Owen, Edwin Asare, Gaia Gullotta, Marta Maria Molinari, and Stefania Sellitti.

The Evaluation Function of the Independent Advisory and Evaluation Service (IAES) extends its sincere gratitude to all participants in the online survey, both within and outside CGIAR, and to those who facilitated its dissemination through communities of practice and beyond.

Special thanks are extended to peer reviewers Ann Marie Castleman and Ahmedou Abdallahi. Additionally, authors are grateful to IAES staff in Rome and consultants who contributed to the achievement of this work, including Agnes Fojta for admin support, Kelly Blank for copy editing, and Federica Bottamedi for layout and publication.

Contents

List of Tables

List of Figures

Table of Acronyms

AfDB	African Development Bank
AR4D	Agricultural Research for Development
CAWI	Computer-Assisted Web Interviewing
CIRAD	Centre de coopération internationale en recherche agronomique pour le développement
EA	Evaluability Assessment
ECB	Evaluation Capacity Building
FAO	Food and Agriculture Organization of the United Nations
GCF	Green Climate Fund
GEF	Global Environment Facility
IAES	Independent Advisory and Evaluation Service
IFAD	International Fund for Agricultural Development
KPI	Key Performance Indicator
MDBs	Multilateral Development Banks
MOPAN	Monitoring Performance Assessment Network
MR	Management Response
MTRs	Mid-Term Reviews
QA	Quality Assurance
R&D	Research and Development
RRA	Responsible Research and Assessment
RRI	Responsible Research and Innovation
SME	Subject Matter Expert
ToC	theory of change
ToR	Terms of Reference
UFE	Utilization-Focused Evaluation
UNEG	United Nations Evaluation Group
UNDP	United Nations Development Programme
WFP	World Food Programme

Executive Summary

Background

While norms and standards have become more harmonized in evaluation, a less standardized but critical part remains: management of the evaluation process. The quality of an independent evaluation is shaped not only by the technical expertise of the evaluation team, but also by the effectiveness of the evaluation's management. Various approaches and modalities are employed by independent evaluation offices within international development agencies. This study explores how evaluation management practices affect use,[1] based on an online survey and literature review focused on the role of evaluation in innovation and strategic impact in Agricultural Research for Development (AR4D). This study maps practices across independent evaluation entities and reveals perceptions of evaluation use in AR4D and multilateral organizations.

This work will help CGIAR build tailored evaluation management arrangements that implement the principles stated in its Evaluation Policy and Evaluation Framework (2022). As per its Terms of Reference (ToR), the Independent Advisory and Evaluation Service (IAES) is the custodian of that policy, liaising with CGIAR governing bodies, presenting proposed revisions for their approval according to best practice and international standards (CGIAR, 2023). This study also aims to spark broader dialogue, within the evaluation community and evidence generation, about the effectiveness of management arrangements in enhancing evaluation relevance and influence.

Methodology

The methodology primarily relied on an online survey to map practices among independent evaluation entities of international development organizations and research institutes; 66 valid responses were collected. In addition, a targeted literature review exploring how evaluation results drive innovation and strategic impact in agricultural research, and a mapping of key features from over 100 evaluations, were conducted. Triangulation of data from different sources and methods was the main analytical approach for developing the conclusions of this study.

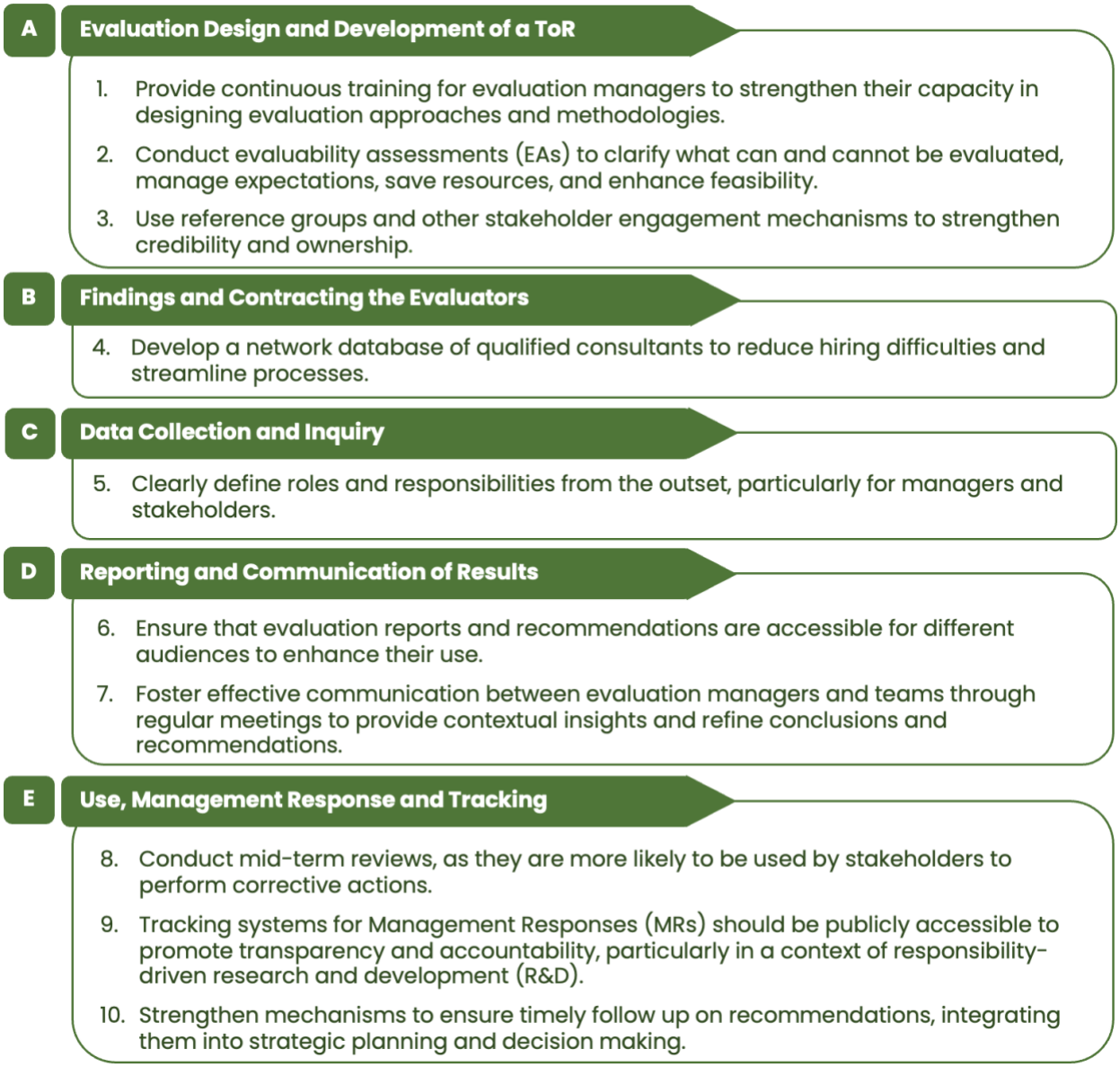
Findings and Recommendations

The management of independent evaluations significantly influences the utilization of evaluation results. The study examined how various international organizations manage independent evaluations, identifying different practices employed across the evaluation process and the key challenges encountered. No direct statistical correlation was established between specific management models and respondents' perceptions of evaluation use. Information gathered from the literature review and evaluation mapping enable us to draw a set of recommendations[2] and conclusions, structured according to the typical phases of an evaluation: (1) Evaluation design and development of ToRs; (2) Findings and contracting the

[1] Evaluation use refers to the ways in which evaluation findings, processes, and recommendations influence decision-making, policies, and actions (M. Q. Patton, 2008).
[2] The recommendations target CGIAR and peer organizations committed to effectively managing independent evaluations and use of evaluative evidence for decision making.

evaluators; (3) Data collection and inquiry; (4) Reporting and communication of results; and (5) Use, management response and tracking.

Figure 1. Key findings and recommendations by evaluation phase



The use of evaluative evidence is primarily a management matter. To foster use, effective management processes should include highly participatory approaches, ensuring that evaluation design, objectives, and scope are tailored to the specific context and available resources. Evaluation managers play a key role in shaping the evaluation methodological framework from the outset and should be well-trained and equipped. Conducting an EA can help save time and manage expectations. Balancing independence and evaluation quality requires a carefully designed and clearly communicated distribution of roles, such arrangement should be adapted to the specific evaluation and context. Mid-term evaluations are more likely to drive course corrections. Finally, tracking systems should be accessible and effectively used to monitor progress and inform decision-making.

1 Introduction and Context

The assessment of international organizations' impact, effectiveness, and efficiency evolved significantly since their inception in the 20th century. Most international organizations have an independent evaluation entity responsible for assessing the organization's contribution to its stated objectives through its programs and projects. Recent decades have seen a growing trend toward greater **standardization of independent evaluations**, driven by an increasing focus on accountability and performance (e.g., OECD-DAC guidelines, UNEG Guidelines), see Annex 3 about principles and standards of independent evaluation in international organizations. The Multilateral Organization Performance Assessment Network (MOPAN) is a network of member countries that fund the multilateral system and share a common interest in enhancing its performance. The assessment follows an evolving generic MOPAN 3.1 indicator framework.[3]

While significant progress has been made in the harmonization of norms and standards, there is a less standardized yet critical dimension that impacts not only the quality of independent evaluations but also its use:[4] Indeed, there is a significant variability in how evaluation offices manage independent evaluations, as shown in several evaluation fora discussions and confirmed by this study. The management of evaluations encompasses various components, including how Terms of References (ToRs)[5] are drafted, the process of hiring evaluators, how stakeholders are engaged, the role of the evaluation manager throughout the process, the timing of report sharing and publication, and the tracking of Management Responses (MRs) and correspondent action plans. Alongside the evaluation team's technical skills, these practices are likely to significantly influence the quality and worth of evaluations.

This study aims to investigate the relationship between evaluation management practices and use through an online survey and through a targeted literature review on how evaluation results drive innovation and strategic impact in agricultural research. It seeks to map practices across independent evaluation entities of peer organizations, while gathering perceptions on the use of evaluations in organizations working on agriculture research for development (AR4D). The findings will be triangulated with existing literature and other analyses.

This investigation is intended to inform CGIAR and other organizations, encouraging reflection on evaluation management and use practices. The 2019 MOPAN assessment of CGIAR identified several weaknesses under Key Performance Indicator (KPI) 8,[6] where CGIAR received a 'highly unsatisfactory'

[3] MOPAN 3.1 indicator framework is organized into five performance areas: Strategic, Operational, Relationship, Performance Management and Results, as well as 12 Key Performance Indicators (KPIs), each with prescribed elements for assessment. KPI 8, under Performance Management, is dedicated to assessing if the organization applies evidence-based planning and programming. It focuses on the evaluation function and its position within the organization's structure, attention to quality, accountability and putting learning into practice.

[4] Evaluation use refers to the ways in which evaluation findings, processes, and recommendations influence decision-making, policies, and actions (M. Q. Patton, 2008).

[5] An essential document in evaluations that defines the objectives, scope, methodology, and requirements, ensuring alignment among stakeholders (OECD, 2010).

[6] 2019 MOPAN assessment of CGIAR identified several weaknesses under KPI 8 about: (1) Accountability and follow-up: lack of a clear accountability system to ensure responses, follow-up, and utilization of evaluation recommendations; and (2) Uptake of lessons and best practices: lack of a formal mechanism for distilling and disseminating lessons

rating. An additional aim is to stimulate discussion in broader evaluation fora, generating further evidence about management arrangements to ensure that evaluation results and recommendations are timely, relevant and influential.

Lastly, this work will help CGIAR build tailored management arrangements to put into practice the principles stated in the Evaluation Policy and Evaluation Framework (2022). As per its ToR, The Independent Advisory and Evaluation Service (IAES) is the custodian of that policy, liaising with CGIAR governing bodies and presenting proposed revisions for their approval according to best practices and international standards (CGIAR, 2023).

1.1 Context of the Study

CGIAR, a global research partnership for a food-secure future, is dedicated to transforming food, land, and water systems in a climate crisis. Operating across various regions worldwide, CGIAR tackles critical challenges in agriculture, food security, and natural resources through diverse research programs and initiatives. CGIAR's evaluation practices are governed by a comprehensive Evaluation Policy and Evaluation Framework (2022) that underscore the importance of independent evaluations in enhancing the quality and impact of its research efforts, as underscored as well by indicators within the MOPAN methodology. Subject to this policy, the IAES Evaluation Function conducts process and performance evaluations that inform strategic decisions and operational improvements. The management within CGIAR systematically tracks recommendations from these independent evaluations, recording how recommendations are addressed and implemented, thus fostering accountability, steering and organizational learning.

In its CGIAR assessment in 2019,[2] MOPAN identified several weaknesses regarding accountability, follow-up, and uptake related to evaluation.[7] Since that assessment, CGIAR underwent multiple strategic and structural changes, resulting in a new Portfolio 2025-30 endorsed in June 2024, and a twice-revised organizational structure. During CGIAR's Portfolio 2022-24, the MR process and products were developed and operationalized. Since its launch in 2019, the Evaluation Function under IAES[8] put significant effort into improving the management of independent evaluations, for example conducting regular After-Action Reviews (a survey sent to an evaluations' key stakeholders to collect their feedback for internal IAES learning). In late 2024, IAES implemented a review of the CGIAR MR system.[9]

learned internally or externally. While some evidence suggests that lessons were applied, the lack of a tracking system hindered assessment uptake.

[7] An official MR to CGIAR MOPAN Assessment was issued, as per MOPAN procedures.

[8] Previously CGIAR Advisory Service Shared Secretariat.

[9] The review aims to help promote the use of evidence from independent evaluations and support evidence-based planning, programming, and decision-making across CGIAR, underpinned by MR System processes. Components of the MR system review are: (1) Inputs (management engagement; recommendations from evaluation teams via IAES, MR template); (2) Process and Outputs (MR development, MR tracking, change management); and (3) Outcomes (implementation status, use of recommendations/ evidence in decision-making). The endorsement process of the MR System Review and its publication on IAES website will occur second quarter 2025.

1.2 Evaluation Results to Drive Innovation and Strategic Impact in Agricultural Research: An Overview of Literature

Chelimsky (1977, 2015) argues that every evaluation responds to a specific demand, which suggests an intention of use. However, the effectiveness of this use depends on factors such as alignment between the evaluation objectives, stakeholders involved, and the quality of the evaluation processes (Patton, 2002, 2008). Chelimsky's **Utilization-Focused Evaluation** (UFE) approach goes further, emphasizing that planning for use must be at the center of the evaluation process. Based on this, independent evaluation management should assume that the evaluation ToR play a fundamental role in establishing from the outset objectives, primary users, and potential applications of the results, thereby ensuring that the evaluation is useful and applied to the context to which it relates directly or indirectly.

Independent evaluation management can act as a catalyst for change, building ToRs that serve purposes ranging from organizational strategy reformulation to guiding investments. On the other hand, as highlighted by Preskill & Boyle (2008, 2009) and Labin et al. (2012), for the results to be effectively used, it is necessary to build a favorable organizational environment that integrates evaluations into strategic planning and promotes a culture of impact. The **feedback loop** to strategic planning is particularly relevant for agricultural research institutions, where evaluation results can be used for research adjustment and feedback (Reed et al., 2021; 2022).

To achieve feedback loops and a culture shift, the process requires not only a clear ToR but also **Evaluation Capacity Building** (ECB), characterized by active collaboration of stakeholders throughout the evaluation process, and institutional analysis of the environment in which the intervention takes place (Better Evaluation, 2023; Cousins et al., 2014; Stockmann et al., 2020, 2022). Thus, use should be seen as a set of intentional actions capable of producing direct or indirect changes, depending on both the planning of the team involved in the evaluation and the organization's senior management (Preskill & Boyle, 2008).

In this regard, Weiss (1979, 1998) and later Alkin & King (2016) **categorize evaluation use into four types**: (1) instrumental use, when findings directly inform decisions and lead to concrete actions; (2) conceptual use, when results enhance theoretical understanding and shape perspectives without immediate application; (3) symbolic use, when findings are used strategically to legitimize pre-existing decisions or positions; and (4) process use, which refers to the learning and organizational changes that occur through engagement in the evaluation process itself.

In research evaluation in general, studies on the use of evaluation results are relatively recent, with room for conceptual and empirical approaches (Milzow et al., 2019; Pinto & Bin, 2024; Van der Most, 2010). For agricultural research and development (R&D) institutions, this discussion began to emerge in the late 1990s, often linked to impact evaluations (Pinto & Bin, 2024).[10] While different types of evaluations co-exist in this field, impact evaluations are particularly prominent in shaping discussions on the use of results. CGIAR is at the forefront of these discussions. The works of Horton & Mackay (2003) and Mackay & Horton (2003) provide a theoretical and reflective analysis of how impact evaluation results should be integrated into the strategies of agricultural research organizations, supporting changes based on organizational learning. They emphasize the need to focus evaluations on practical use to maximize their relevance and

[10] A systematic process to determine the changes attributable to an intervention or program, focusing on both intended and unintended outcomes.

usefulness in decision-making. This perspective is later expanded upon by Hall et al. (2003) and Patton & Horton (2009).

Hall et al. (2003) criticize the narrow focus of impact evaluations on economic outcomes. They advocate for a **broader approach grounded in the concept of innovation systems**, which incorporates a more comprehensive framework designed to foster collaboration, institutional learning, and systemic innovation. This broader scope aims to provide R&D managers with an integrated and holistic view of the innovation process, ensuring that impact evaluations reflect the institutional context of agricultural research interventions. Hall et al. emphasize that the responsibility for using evaluation results should be shared among multiple actors, making the use of results a collective effort, particularly within innovation systems. More recently, SPIA (2020) highlights that CGIAR integrates impact evaluations not only for accountability purposes but also into resource allocation and research prioritization, reinforcing their role in shaping institutional learning and guiding future investment. In this perspective, all good research contributes to knowledge, but only some of that knowledge leads to insights or innovations that can be scaled and contribute to real-world impacts. While AR4D relies on specific theories of change (ToC) to link research to impacts at scale, the ToC necessarily makes many assumptions along the long pathways to impact (SPIA, 2020).

Patton & Horton (2009) adopt a different approach, emphasizing that the use of **evaluation results are closely linked to the role of the evaluators.** They propose a comprehensive model that ensures evaluations are use-centered, focusing on identifying key users, ensuring their engagement, and effectively communicating results. The model also stresses the importance of building evaluation capacity within organizations and fostering stakeholder involvement throughout the process.

This approach is guided by an **Adaptive Cycle**, which includes proactive actions, adjustments based on feedback, continuous interaction with users, and adaptation as needs evolve. These phases aim to make the evaluation process reflexive, allowing evaluators to adjust focus and methods as stakeholder needs become clearer. To maintain integrity and credibility, Patton & Horton (2009) highlight the need for a balance between active user participation and adherence to rigorous quality standards, ensuring that evaluations are impartial and highlight both strengths and weaknesses of the program.

Joly et al. (2016) studied five agricultural R&D organizations with the objective of gathering information that would contribute to improving the ASIRPA method (Analyse de l'impact sociétal de la Recherche). During the analysis, they found that the use of evaluation results for accountability, followed by advocacy, predominated overuse for organizational learning, including R&D management. They emphasize the importance of using evaluation information to improve research and maximize its impact, stressing the continuous use of results throughout the entire research cycle, as proposed in ASIRPA.

Pinto & Bin (2024) recently conducted a study with eight agricultural R&D organizations to analyze how impact evaluation results are used. The authors applied the **4A's evaluation framework**, proposed by Morgan et al. (2013; 2017), as shown in Table 1, and found that the use of impact evaluations for learning surpassed its use for advocacy.

Table 1. 4A's evaluation framework

A's of Assessment	Description
Accountability	To demonstrate that money and other resources were used efficiently and effectively, and to hold stakeholders accountable.
Analysis	To understand why, how, and if the research is effective, and how it can be better supported.

Advocacy	To demonstrate the benefits of supporting research and to improve understanding of research and its processes among policymakers and the public.
Allocation	To determine how to distribute funding across the research system.

Pinto & Bin demonstrate that, in these institutions, the use of results is mainly concentrated on Accountability (reporting to funders), with seven institutions applying the results for this purpose, which is similar to the findings of Joly et al. (2016). Notably, in CGIAR and other development assistance contexts, there is a growing emphasis on directing impact assessments toward accountability, particularly in terms of return on investment, cost-effectiveness, and cost-benefit analyses. Five institutions used the results for Analysis (organizational learning), four for Advocacy (demonstrating value to society), and two for Resource Allocation (informing resource distribution). Moreover, Pinto & Bin highlight that the use of results has evolved over time to become a transformational element, contributing to impacts across different dimensions, whether economic, social, environmental, or institutional.

Prioritization of uses for Analysis and Advocacy demonstrates that agricultural research impact evaluation is a process beyond measuring changes that have occurred or may occur against a certain investment. Evaluation thereby becomes a support tool for agricultural innovation, integrating with movements that demand responsible research and evaluation, such as Responsible Research and Innovation (RRI) and Responsible Research Assessment (RRA). These movements aim for research that is ethical, transparent, and socially impactful (Schönbrodt et al., 2022; Schuijff & Dijkstra, 2020). The use of evaluation results by institutions—whether for Accountability, Analysis, Advocacy, or Allocation-can play a key role in guiding them toward greater impact. As demonstrated by Morgan et al. (2017) and Pinto & Bin (2024), structured and intentional use of evaluation results can foster institutions to achieve more impact. This aligns with the broader premise of RRI and RRA.

1.2.1 Findings on Evaluation Use in Agricultural R&D

Pinto & Bin (2024) identified factors that facilitate or hinder the use of evaluation results in agricultural research, grouping them into three categories. In their investigation, authors show that communication of agricultural R&D evaluation results, often seen as the crucial element for ensuring use, is one of these factors, but is also linked to others:

- **Category 1: Structural and Organizational Factors**: Support, resources, and strategic relevance of evaluations.
- **Category 2: Operational Factors**: Quality, rigor, appropriate methods, and timely communication.
- **Category 3: Applicability Factors**: Literacy in evaluation processes, stakeholder pressures, and credibility of findings.

Authors identify a clear gap in the management and systematization of impact evaluation results. None of the eight institutions had an established process for the use of the results, nor a system to record MRs[11] or feedback. The growing global emphasis on RRA, RRI, responsible investment, and mission-oriented

[11] The MR provides management's views of the evaluation recommendations, including whether and why management agrees or disagrees with each recommendation. The MR should detail specific actions to implement those recommendations that were agreed to by management. These actions should be concrete, objectively verifiable, time-bound and clear on the responsibilities for implementation. (UNEG, 2016).

research, combined with the Sustainable Development Goals (SDGs), reinforces the importance of evaluation as a tool for social transformation (Von Schomberg, 2019). Establishing an organizational culture focused on societal impact means incorporating evaluation not just as a bureaucratic requirement, but as a catalyst for change. Pinto & Bin (2024) highlight that agricultural R&D institutions can use evaluation results to recalibrate research focus, optimize project design, and influence resource allocation, which would support socially beneficial innovations and promote positive impacts across multiple dimensions.

This perspective aligns with the concept of '**Impact Culture'**, in which agricultural research evaluations are continuously used to guide research towards societal impacts, from its inception to its conclusion and beyond (Ferre et al., 2023; Ferré et al., 2025). In this culture, evaluation becomes a dynamic tool that guides decisions at all stages of the research process, promoting continuous learning and improvement. Therefore, the use of results is an integrated activity and a constant process, guiding research to demonstrate actions aimed at transforming society.

This approach is also connected to **Transformative Evaluation** (Mertens, 2009), which emphasizes the role of evaluations in fostering social change within a framework of responsibility. Within this paradigm, the use of impact evaluation results can complement and reinforce other types of evaluations, such as performance and process evaluations, guiding agricultural R&D institutions toward an impact-oriented culture. By integrating evaluation findings into institutional decision-making, these organizations can align their planning processes with societal transformation goals, using evaluations as strategic tools for analysis, monitoring, and guidance. Such an integrated approach resonates with the ethical principles of RRI and RRA.

1.3 Study Purpose and Scope

Through the study and associated reviews on evaluation management styles, the IAES Evaluation Function of CGIAR aimed to gain an overview of evaluation management practices in international organizations. The study maps these practices across independent evaluation entities of peer organizations, while gathering perceptions on the use of evaluations. This study has a dual purpose:

- Advance the state of the art of evaluation management practices, particularly in organizations implementing research and AR4D.
- Directly contribute to planned 2026 review of CGIAR Evaluation Policy and Framework.

The scope is understanding evaluation management practices in UN agencies, international and regional development banks, donors, and other relevant organizations, particularly in organizations implementing research and AR4D. This study will support IAES in defining its own best practices and aligning them with widely recognized and approved norms in development evaluation. The findings are expected to enhance the use of evaluative evidence in decision-making processes.

1.4 Methodology and Data

This study methodology primarily relied on an online survey to map practices among independent evaluation entities of international development organizations and research institutes, a targeted literature review exploring how evaluation results drive innovation and strategic impact in agricultural research, and a mapping of key features from over 100 evaluations. Findings were cross-checked with existing literature

and other analyses, for example the EvalforEarth online discussion, to enhance validity. Triangulation of data from different sources and methods was the main analytical approach for developing the conclusions of this study.

The survey targeted professionals with experience in managing independent evaluations within international organizations focused on international development and research. These included UN agencies, international and regional development banks, donors, and other relevant organizations. It was advertised through global and regional evaluation networks and associations, such as EvalforEarth,[12] Peregrine Discussion Group,[13] and EvalMena,[14] and responses were collected via Computer-Assisted Web Interviewing (CAWI). Participants provided their responses autonomously and anonymously from 20 August-7 November 2024.

The **online survey** explored key aspects of managing independent evaluations and the level of involvement required from evaluation managers, officers, and specialists. It was structured around seven main topics related to evaluation management: (1) Types of evaluations conducted; (2) Drafting the ToR evaluations; (3) Hiring evaluators; (4) Data collection; (5) Report writing; (6) Publication and use of evaluations; and (7) MR.

For recruitment, an approach related to convenience sampling was chosen following these criteria:

1. **Voluntary participation via professional networks:** Only those who responded to the initial call through posts on networks such as EvalforEarth, Peregrine Discussion Group and EvalMena.
2. **Autonomy in response:** Participants spontaneously joined the survey.
3. **Relevant professional engagement:** Only participants with experience in independent evaluation in agricultural R&D institutions or organizations related to international research and development were included.
4. **Expanding the sample (in a complementary way**): Some participants were reached through organic sharing, following the snowball method (Parker et al., 2019), but this was not the central method of the research.

The survey received a total of 84 responses, which proceeded to data cleaning. Among respondents, six people did not qualify as 'People with experience in independent evaluation management', while another five declared as having experience but later in the survey clarified that they were either independent evaluators or employees of a consulting company. Seven qualified as evaluation managers but did not provide an answer to the survey beyond the first question on whether they had experience in this field. The final number of valid responses amounts to 66.

The purpose of the **desk review mapping** was to gather insight into the output of evaluation functions of international organizations with missions similar to CGIAR. In this desk review, target organizations were identified, which were the Rome-Based UN Agencies, Development Banks, and other organizations whose missions and themes align with CGIAR. For each organization a sample of their recent evaluation reports were analyzed against a few characteristics, including which types of evaluations are being conducted, the time required to publish reports, the number of countries visited, and the size of the teams involved in each evaluation. Nine external organizations were covered, and the selection of reports to be analyzed ensured

[12] EvalforEarth is a Community of Practice on Evaluation for Food Security, Agriculture and Rural Development. Website accessed 1/17/2025: https://www.evalforearth.org

[13] A community of practice managed by IOCE/EvalPartners. Website accessed 1/17/2025: https://evalpartners.community/peregrine

[14] The Middle East and North Africa Evaluation Network. Website accessed 1/17/2025: http://www.evalmena.org/

the coverage of all types of evaluation reports published by the organizations.[15] One or more report for each theme was selected randomly among the publicly available ones.[16] Two evaluation reports from CGIAR IAES were also included, reaching the final number **of 100 analyzed evaluations**. Distribution of the evaluation reports considered is as shown in Table A1 (Annex 4). Reports were downloaded from websites of the evaluation functions of organizations at issue in **November 2023 and August 2024**. All results of the mapping exercise are in Annex 4.

2 Results of the Online Survey

2.1 Respondents' Profile

As shown in Figure 2 (a) and (b), survey respondents were mostly female (53%), with males accounting for 44% and 3% preferring not to respond. Almost all participants were over the age of 30, with the largest demographic being those aged 41-50 years old (39%). This was followed by respondents aged 31-40 years (27%), those over 61 years (17%) and individuals aged 51-60 years (14%).

Responses were received from across all regions, providing a diverse range of perspectives, though not equally represented. The variance in the representation may be associated with the location of the evaluation offices of the agencies responding to the survey. Most participants (55%) were based in Europe, Latin America and the Caribbean accounted for 18%, followed by 9% from Sub-Saharan Africa. A further 17% was equally distributed across South Asia, Middle East and North Africa and North America, the detailed distribution is shown in Figure 2 (c).

Figure 3 depicts the distribution of respondents based on the type of organization they work for. Over one-third (36%) are employed by a UN agency, making it the largest group represented. This is followed by individuals working for government entities, who account for 27% of the respondents, those affiliated with international research organizations make up 15%, while both donor organizations and implementing organizations each contribute 8% of the total. Additionally, 5% of respondents are from other multilateral organizations or funds, and a smaller group (2%) are employed by development banks.

As shown in Figure 4, respondents of the survey reflect a mix of professionals at different points in their careers. Most respondents are mid-to senior evaluation managers, as over 60% have more than eight years of experience, including a significant 14% with over 20 years of experience. Additionally, 27% of respondents have between four and seven years of experience, while 11% are junior managers with zero to three years of experience.

Finally, regarding how respondents allocate their time to evaluation management, Figure 5 illustrates that for 30% of participants, evaluation management accounts for more than 75% of their work time. A further 15% dedicate at least half of their time to this work. Meanwhile, 17% of individuals spend between 30% and

[15] Each organization has a specific system for categorizing evaluation reports. At least one report from each category that the organization uses were sampled. After the selection based on organization-specific categorization, categories across the nine organizations were harmonized. (see Table A2 of Annex 4 for more details).
[16] In each organization, all the reports published between 2018 and 2023 were listed, classified by category. At least one report was randomly selected for each category, for a total of ten to twevle reports for each organization. More details are available in Annex 4.

49% of their time managing evaluations, while one third of respondents manage evaluations on a less intensive basis, devoting less than 30% of their time to it.

Figure 2. (a) Age, (b) gender and (c) region of respondents (N-66)

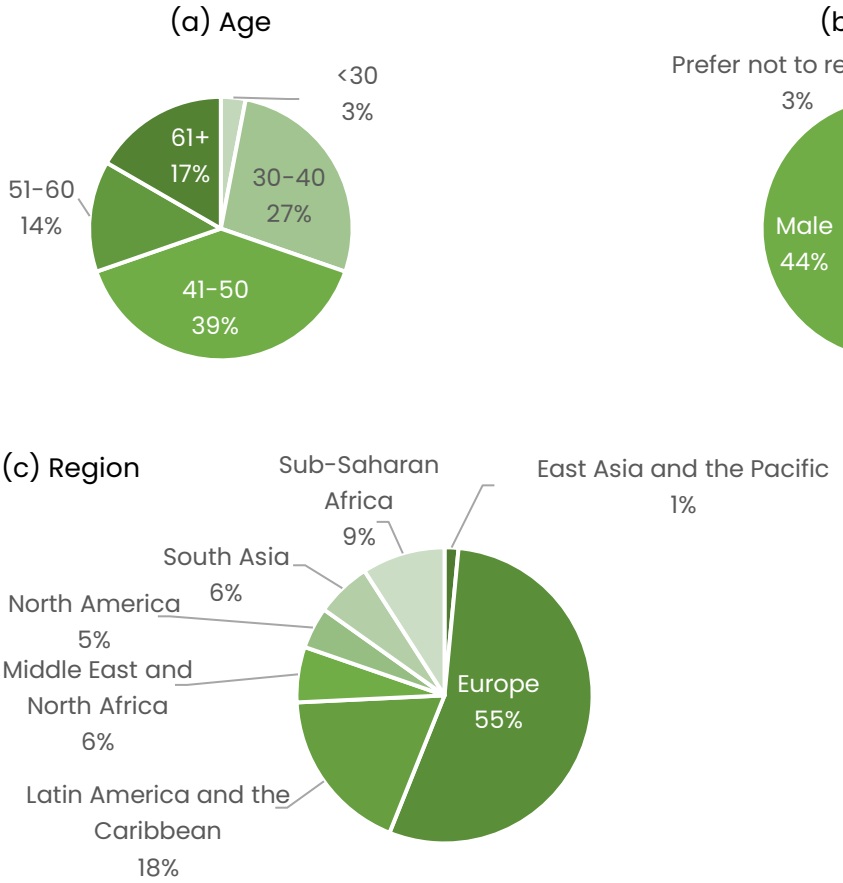


Figure 3. Distribution of respondents by type of organization (N-66)

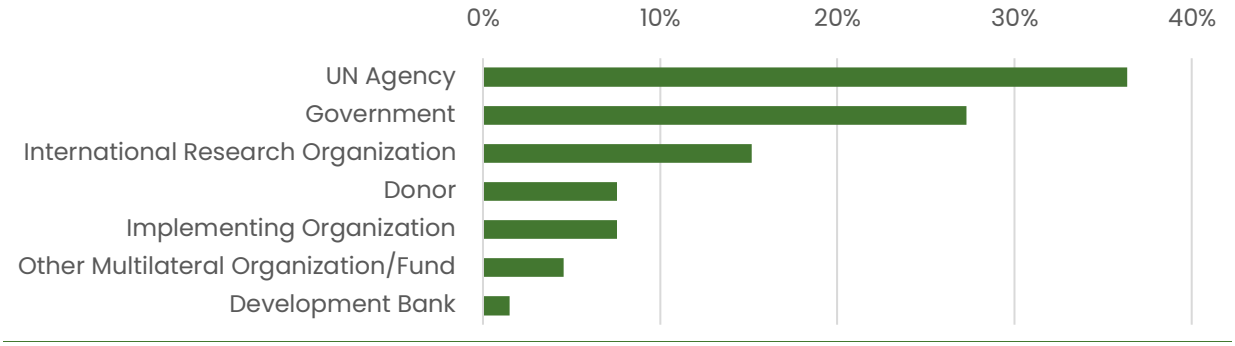


Figure 4. Distribution of respondents by years of experience managing evaluations (N-66)

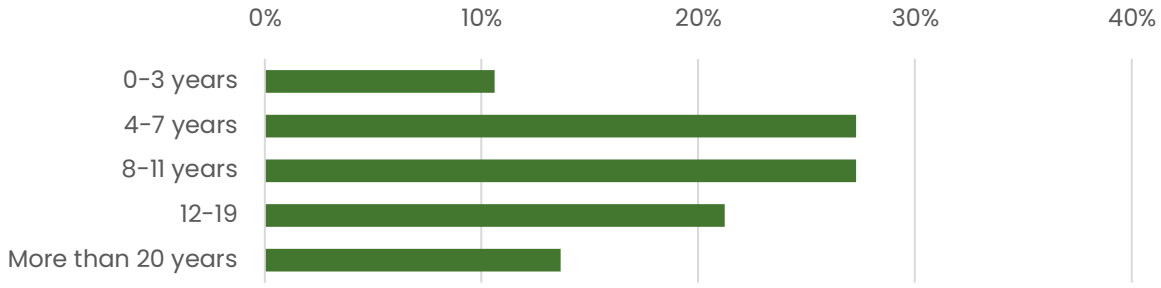
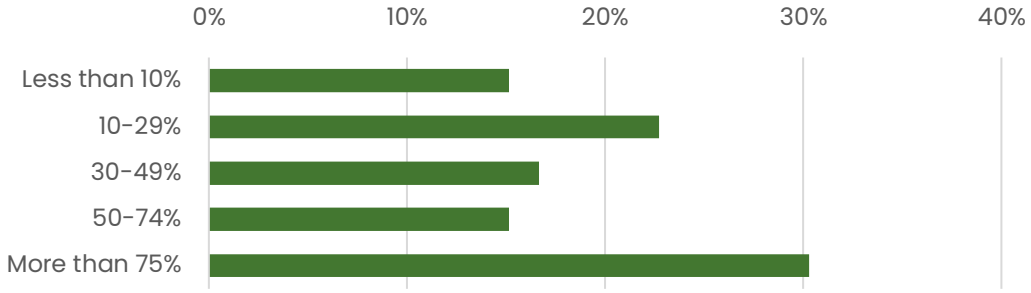


Figure 5. Distribution of respondents by time allocated to evaluation management (N-66)



2.2 Types of Evaluations

Figure 6 highlights the types of evaluations that respondents typically manage. Project and program evaluation emerge as the most frequently managed type of evaluation, managed by 82% of respondents. Approximately half manage thematic, cluster or sector evaluations, as well as regional or country-level evaluations and impact assessments. Around 30% of respondents manage corporate evaluations, as well as synthesis, and reviews or stocktaking. Some respondents mentioned other types, most notably socio-environmental assessments.

As detailed in Table 2, the distribution of evaluation types managed varies by organization.[17] Respondents from UN agencies, government entities, and international research organizations oversee all listed types of evaluations. Participants from development banks and other multilateral organization funds also handle a broad spectrum of evaluations, except for reviews and stocktaking for the former, and synthesis and impact assessments for the latter.

Donor and implementing organizations manage a more specialized range of evaluations. Donor organization respondents primarily handle project and program evaluations, thematic, cluster, or sector evaluations impact assessments and regional and country-level evaluations. Implementing organizations, as expected, focus mainly on project or program evaluations, impact assessments, and reviews or stocktaking.

[17] Note that respondents were not required to indicate the exact organization they worked for, but only the type of organization. Even though organizations may cover different roles at different times, for instance the donor as well as the UN Agency, the role reflected by respondents was followed when asked about the type of organization.

Figure 6. Share of respondents usually managing each type of evaluation (N-56)

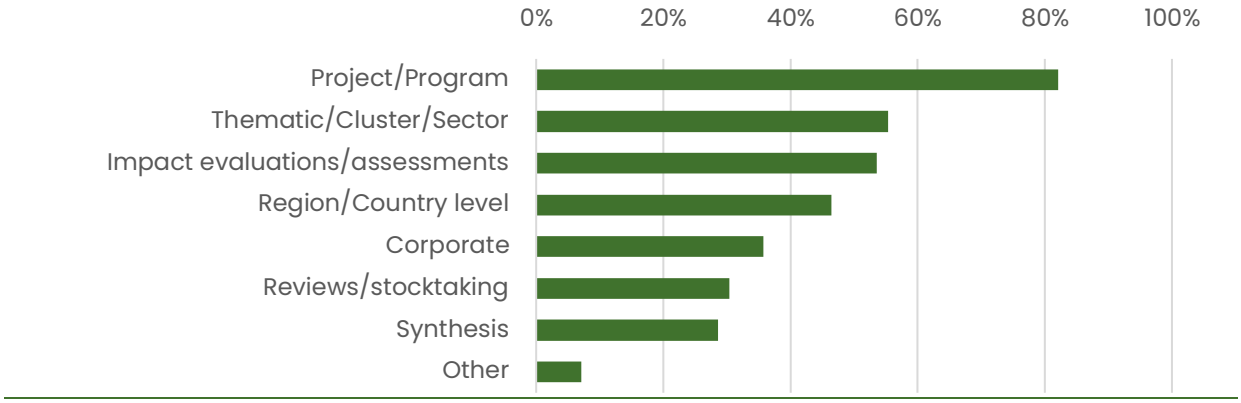


Table 2. Type of evaluations managed by respondents

Type of organization	Thematic / Cluster / Sector	Project / Program	Corporate	Region / Country level	Synthesis	Impact evaluations / assessments	Reviews / stocktaking	Other
UN Agency	✓	✓	✓	✓	✓	✓	✓	
Government	✓	✓	✓	✓	✓	✓	✓	✓
International Research Organization	✓	✓	✓	✓	✓	✓	✓	✓
Donor	✓	✓		✓		✓		
Implementing Organization		✓				✓	✓	
Other Multilateral Organization/Fund	✓	✓	✓	✓			✓	
Development Bank	✓	✓	✓	✓	✓	✓		

2.3 Terms of Reference

This section of the survey focused on the initial phase of managing an independent evaluation: drafting the ToR. Overall,[18] 77% of respondents declare having led the development of ToRs for an independent evaluation. All respondents from UN agencies, donor organizations and other multilateral organizations

[18] From this point forward, all "total" figures also include responses from the single respondent representing a development bank. However, this response has not been highlighted as a separate category.

and funds were responsible for this task. In contrast, this share drops to 80% for international organizations, 67% for implementing organizations (Figure 7) and ess than 40% for respondents from government entities.

More than 73% of respondents indicate that developing a ToR takes less than ten days (see Figure 8). Respondents from donor organizations and implementing organizations report a quicker process, with many indicating that it takes a maximum of five days. Longer ToR drafting processes are reported by individuals from UN agencies, government entities, and other multilateral organizations and funds, where most respondents indicate drafting times over six days. Individuals from international research organizations are equally distributed between those that indicate less than a week and those that report longer times.

Seven of ten (70%) respondents indicated that **the evaluation manager** is primarily responsible for the design of the evaluation approach, methodology and formulation of the main questions. It was followed by 15% the consultant or firm, 9% another person and 6% the entity that commissioned the evaluation (Figure 9).

Figure 10 presents an overview of additional individuals or groups involved in the evaluation design process, beyond the primary responsible parties identified earlier, categorized by type of organization. The evaluand team is cited as a key participant by most respondents across all organizations, except for donors. Donors, in contrast, primarily identify the entity that commissioned the evaluation as the key additional figure involved.

Respondents from UN agencies most commonly report the involvement of the evaluand's main stakeholder in the design process. Consultants or firms are consistently involved across all organizations. Other individuals mentioned in the comments as being part of the process include steering committees and evaluation supervisors.

Figure 7. Have you been in charge of developing the evaluation ToRs? (N-60)

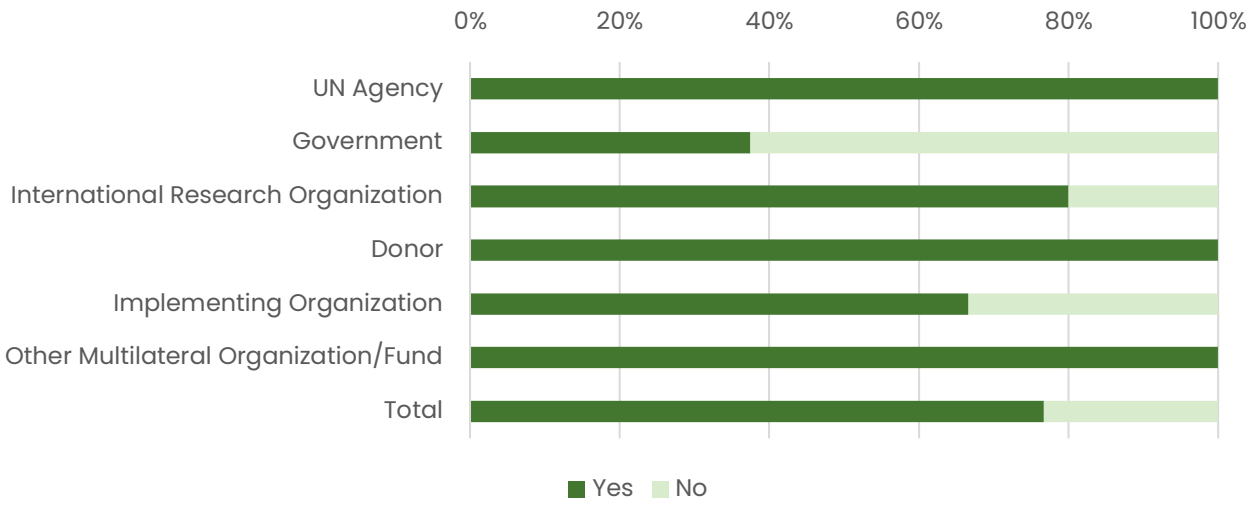


Figure 8. Usual time spent drafting ToRs for evaluations (N-47)

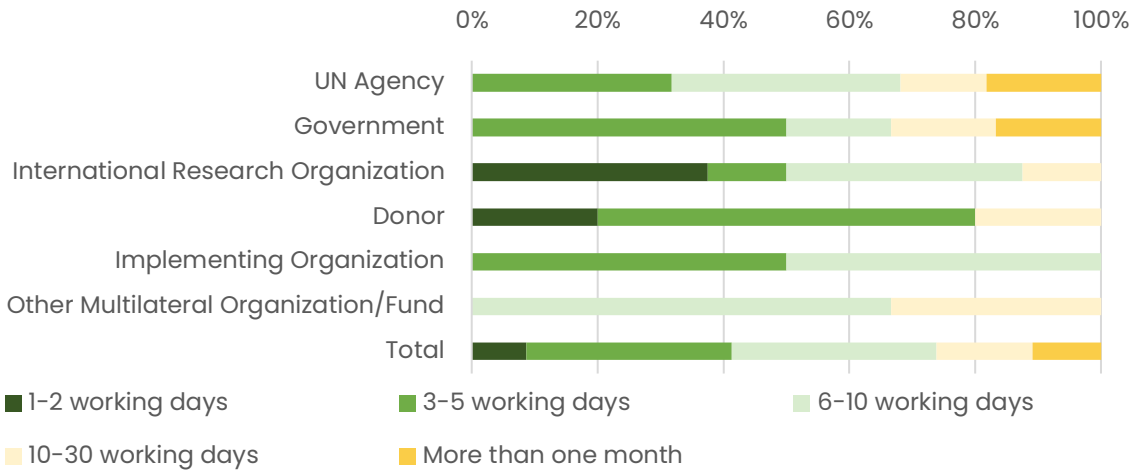


Figure 9. Primary responsible for the design of the evaluation approach, methodology and the formulation of main questions (N-47)

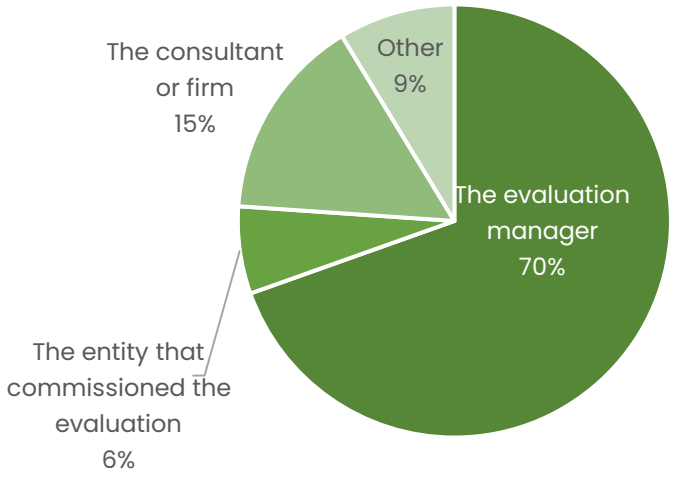
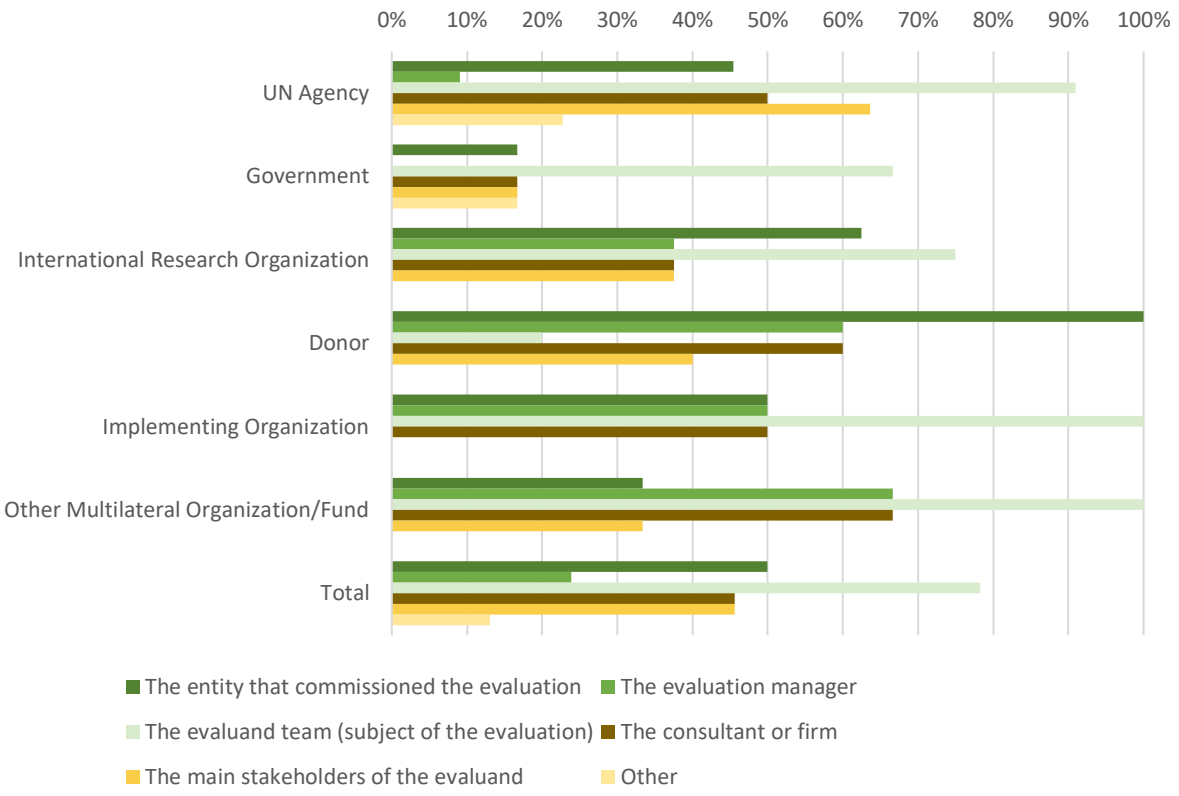


Figure 10. Who else participates to/formulates the evaluation questions? (N-46)



In the survey, information about **evaluability assessments (EAs)** were given,[19] e.g., whether they are usually carried out, and if so, when in the evaluation process they are carried out, and whether respondents had experience conducting them.

Overall, 35% of respondents indicated that such assessments are conducted either consistently or most of the time, while the majority reported them occurring more sporadically (Figure 11). According to respondents, these assessments are more consistently carried out in international research organizations, UN agencies and implementing organizations, followed by donors and government entities. EAs are mostly reported as a rare occurrence in other multilateral organizations and funds.

More than half (about 60%) of respondents report that EAs are carried out either before the evaluation or during the evaluand design (see Figure 12). Another 20% report this activity as occurring either during the evaluand implementation or the evaluation inception phase. EAs are carried out before the evaluation or during the evaluation design for most respondents from UN agencies, international research organizations, implementing organizations and other multilateral organizations and funds. Some respondents selected the 'other' option and elaborated in the comments that the timing of EAs depends on the context. For topics under exploration, they may occur earlier, while for other evaluations, they may happen during the design phase if deemed necessary. Some related work is conducted during work planning or risk

[19] A preliminary determination of whether a program or intervention has sufficient clarity in its objectives and available data to be evaluated (Wholey, 2004). To learn more about evaluability assessments visit the IAES online portal: Evaluability Assessments: Enhancing Pathway to Impact

assessments, and others during the ToR or evaluation inception phase; however, comprehensive EAs are rarely undertaken beforehand. For project or program evaluations, such assessments are often unnecessary due to the use of standard approaches. In some cases, a lighter version of the process, such as during an intake or work planning stage, is used.

More than half of respondents managed an evaluation with an EA (Figure 13). This figure increases to over 70% of respondents from UN agencies, about 60% of donors and other multilateral organizations and funds.

Figure 11. Is an EA usually carried out? (N-60)

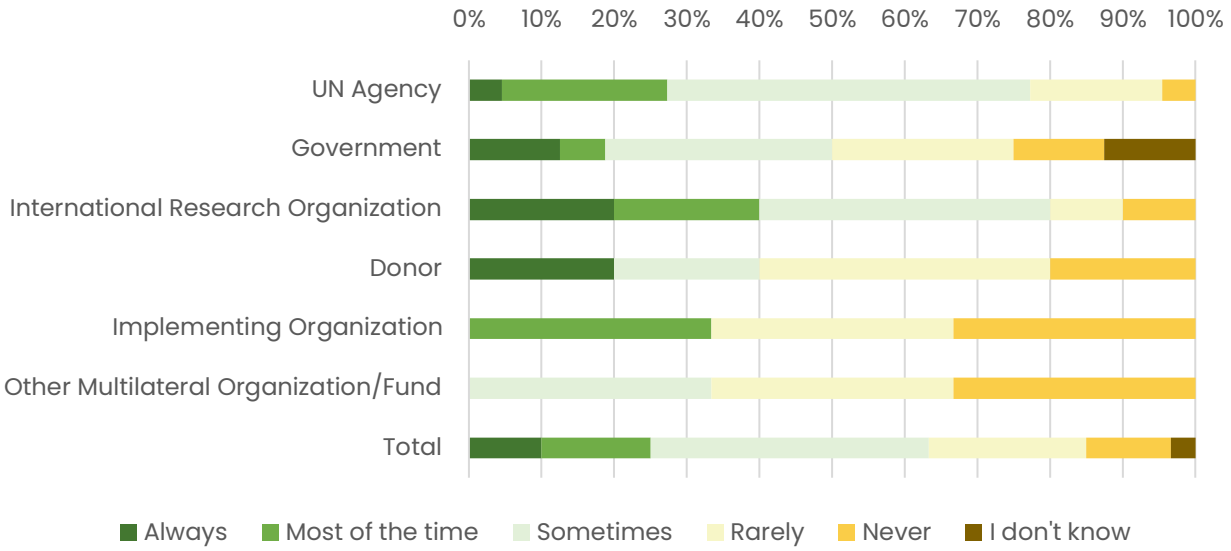


Figure 12. When is the EA usually carried out? (N-58)

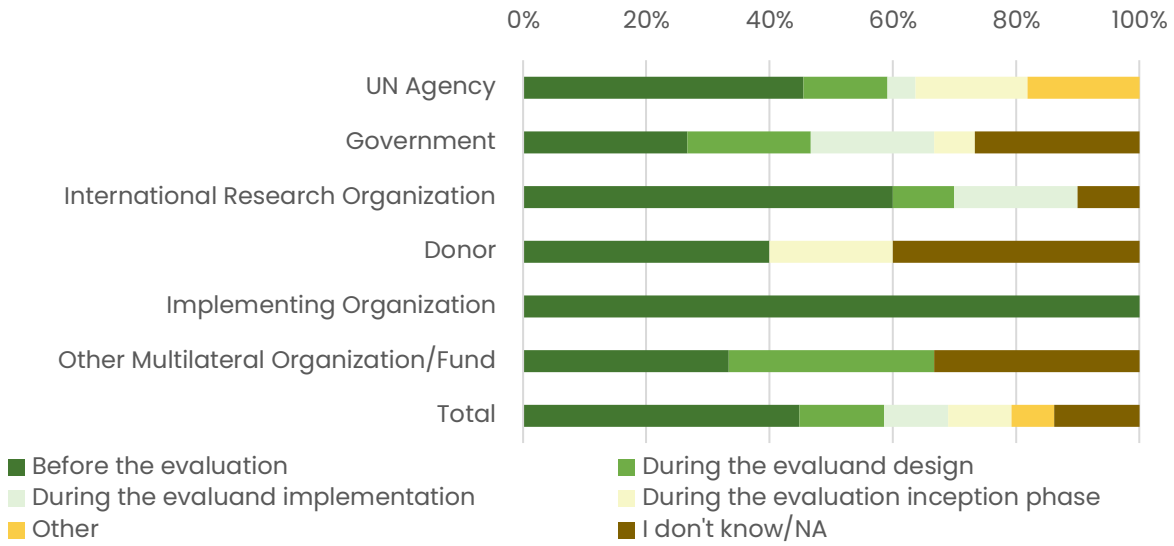
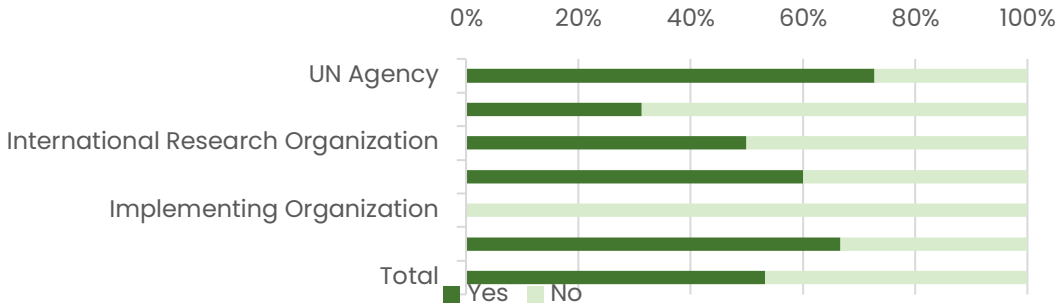


Figure 13. Have you ever managed an evaluation with an EA? (N-60)



2.4 Contracting the External Evaluators

This section of the survey explored challenges related to **hiring evaluators**, including preferences for individual consultants versus firms, difficulties in assembling the right team, the time required for the hiring process, and overall satisfaction levels among respondents.

Figure 14 provides a breakdown of the respondents' preferences for individual consultants or firm by type of organization. Respondents from UN agencies have a strong preference for individual consultants, while the other organizations do not show such clear preference and highlight that it depends on the type of evaluation, evaluand and context. Individual consultants have the advantages of an easier contracting process, lower costs and provide the highly specific expertise that is sometimes required. They are often preferred for project-level evaluation, which typically have lower budgets. Firms, on the other hand, can be deemed to provide quality assurance (QA), backstopping, and credibility; regulatory frameworks may favor firms, and they may be more suited for complex or large-scale evaluations. Other approaches mentioned include cases in which firms are hired, but the work is carried out by specific consultants, and cases in which evaluations are conducted in-house without external consultants.

Most respondents indicated that the hiring process typically takes less than one month (Figure 15). Satisfaction levels of respondents with hiring consultants or firms are positive across all organizations, with most respondents reporting being very satisfied or satisfied. A smaller proportion feel neutral about it, and only a few reported some dissatisfaction (Figure 16). Despite this, 60% of respondents considered finding and hiring the right team to be somewhat difficult or very difficult (Figure 17). Only a small number of respondents from UN agencies and international research organizations described the process as very easy, while the majority either expressed neutrality or encountered some level of difficulty.

Figure 18 provides an overview of the challenges most frequently ranked among the top three when hiring an evaluation team. Time constraints emerge as the most cited challenge (29%), followed equally by the low availability of subject matter experts (21%), budget constraints (21%) and long bureaucratic processes (21%). Only 8% of respondents identified their lack of knowledge of the local context and local consultants as a top three challenge. Additional insights from the comments expand on these challenges, highlighting the skill gap among subject matter experts (SMEs) who may lack experience in evaluation processes. Respondents noted the difficulty of finding consultants with both subject matter expertise and evaluation knowledge, combined with strong analytical and writing skills. Another significant challenge relates to managing individuals, which can be time-consuming. Coupled with challenging team dynamics, this can negatively impact the quality of the evaluation.

Figure 14. Do you mainly hire firms or individual consultants to conduct independent evaluations? (N-54)

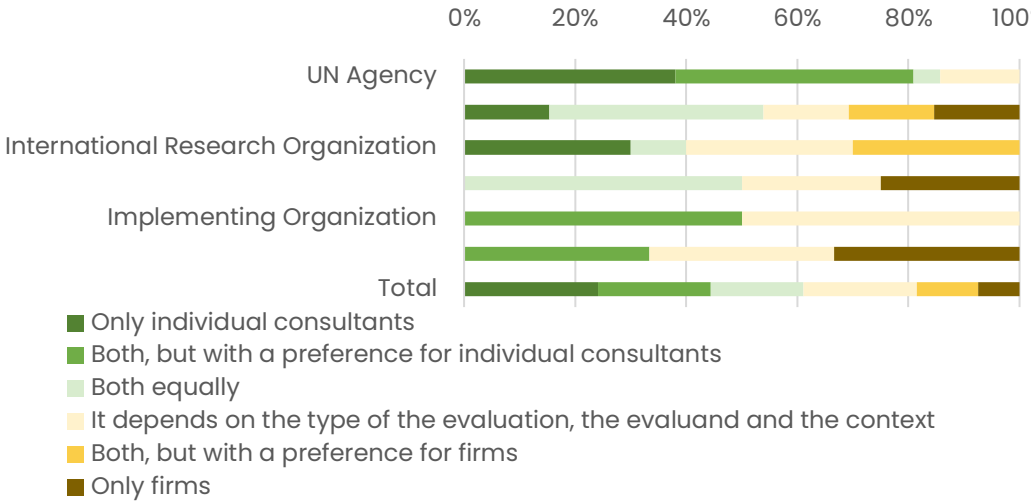


Figure 15. Time spent finding the right team (N-32)

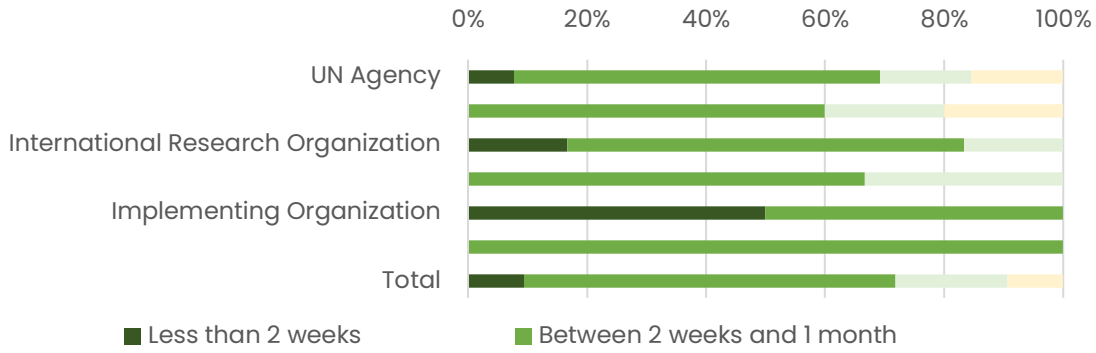


Figure 16. Level of satisfaction with hiring consultants or firms (N-35)

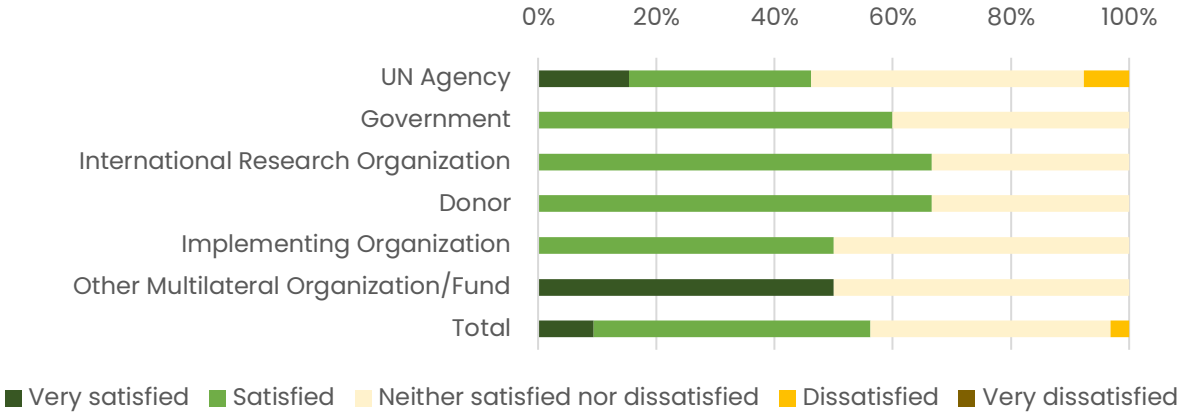


Figure 17. Level of difficulty of finding and hiring the right team (N-35)

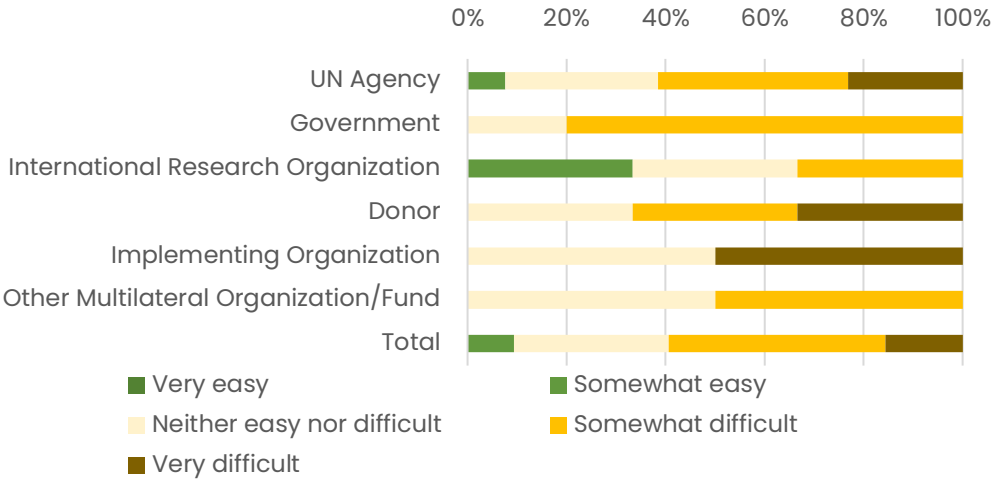
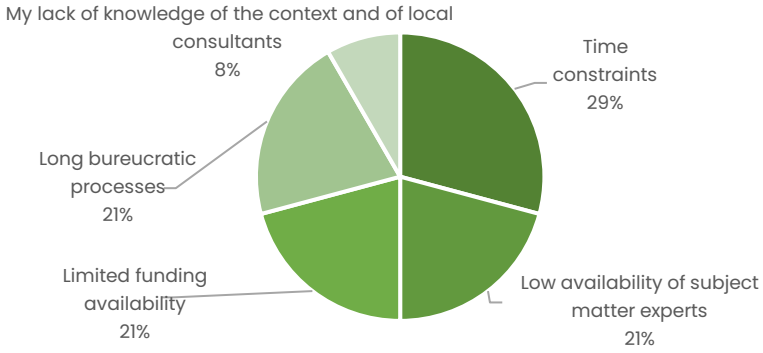


Figure 18. Which are the top three challenges in finding the right individual consultant/team of consultants/firms (N-32)



2.5 Data Collection

This section explored evaluation management dynamics related to data collection for evaluation purposes. Findings reveal that virtually all respondents are involved in the design phase of data collection. Respondents from other multilateral organizations and funds, stood out as an exception to this trend, with approximately half reported being rarely involved in this process (Figure 19).

Regarding field work, the sample is evenly split between those who do so either always or most of time and those rarely or never participating. Respondents from donor organizations, implementing organizations and other multilateral organizations or funds rarely travel for fieldwork. In contrast, those from UN agencies, government entities and international research organizations participate more frequently in field missions. Specifically, about half of respondents from UN agencies and international research organizations, and about 30% of government entities respondents report participating in field work (Figure 20).

Respondents' involvement includes participating in interviews, focus groups and other data collection activities, with 67% of respondents reporting engagement in this area (Figure 20). About half of survey participants stated they are actively involved in asking questions during data collection activities rather than participating solely as observants. This is particularly noticeable among respondents from UN agencies, government entities, international research organizations and donor organizations (Figure 21). Evaluation manager involvement in interviews varies by context. Consultants usually take the lead, but managers step in when necessary to clarify, refocus, or address unanswered questions. In some cases, the manager's role is outlined in the ToRs, requiring them to take a more active role, particularly when acting as evaluators or having a direct link to the interviewee. Overall, their participation depends on the evaluation's specifics and team dynamics.

Figure 19. In your role, do you contribute to the data collection design? (N-43)

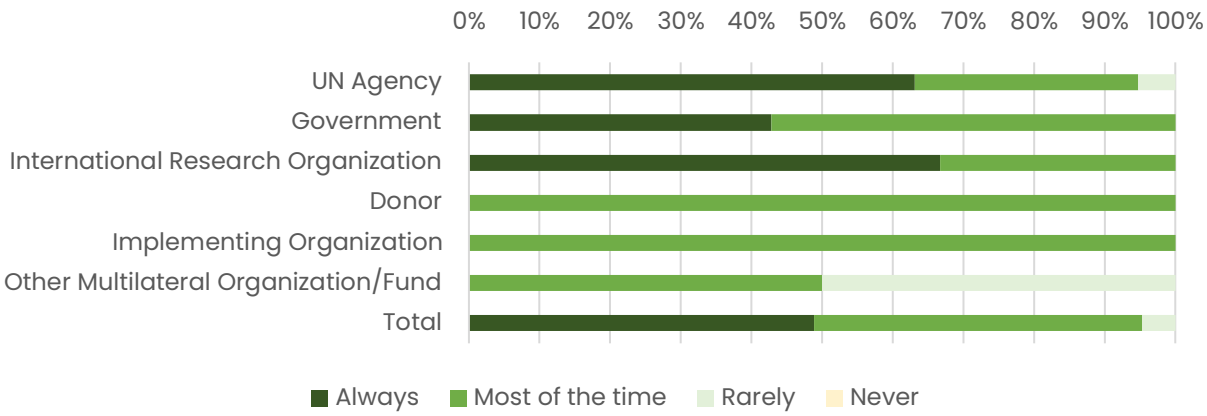


Figure 20. Do you travel to the field during evaluations? (N-43)

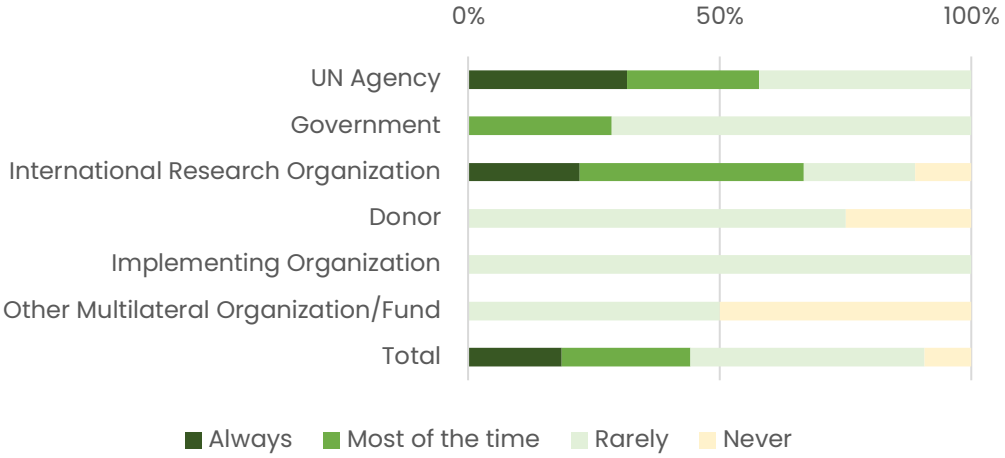


Figure 21. Do you participate in interviews, focus groups and other data collection activities? (N-43)

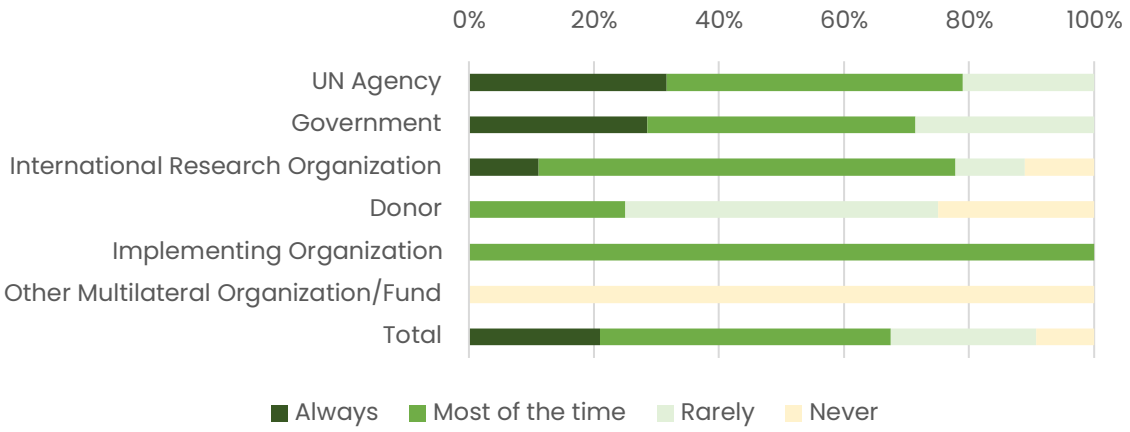
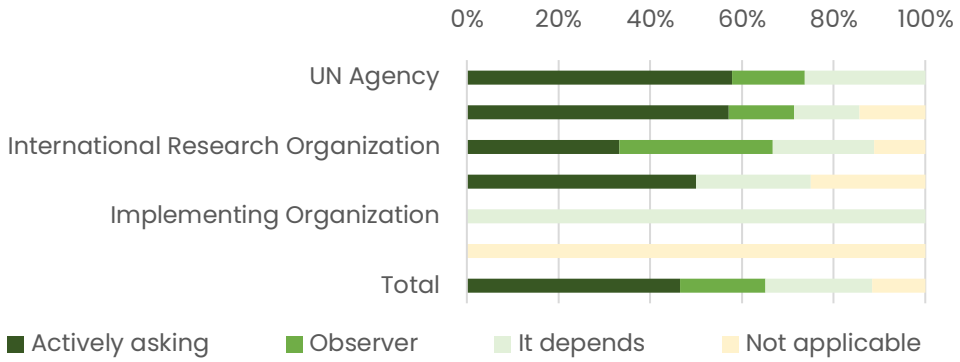


Figure 22. Do you participate as an observer, or do you actively ask questions? (N-43)



Respondents were asked an open-ended question on the **pros and cons of participating in data collection as evaluation managers**. Figures 23 and 24 present word clouds of the most frequently mentioned words associated with the advantages and disadvantages of direct participation.

Many respondents viewed **participation** to **improve the overall quality** of evaluations. They emphasized benefits such as enhanced data accuracy and robustness, quicker identification and resolution of issues in data collection tools and acting as an 'insurance policy' to uphold evaluation quality in case consultants underperform. Additionally, participation was seen as fostering **credibility with evaluands** and stakeholders, building trust, and enhancing engagement. Respondents also highlighted that direct involvement promotes **ownership of findings**, leading to higher-quality evaluations and actionable recommendations. Furthermore, it allows a more **nuanced understanding** of the evaluand and the evaluation process, aiding in evidence triangulation and improving transparency.

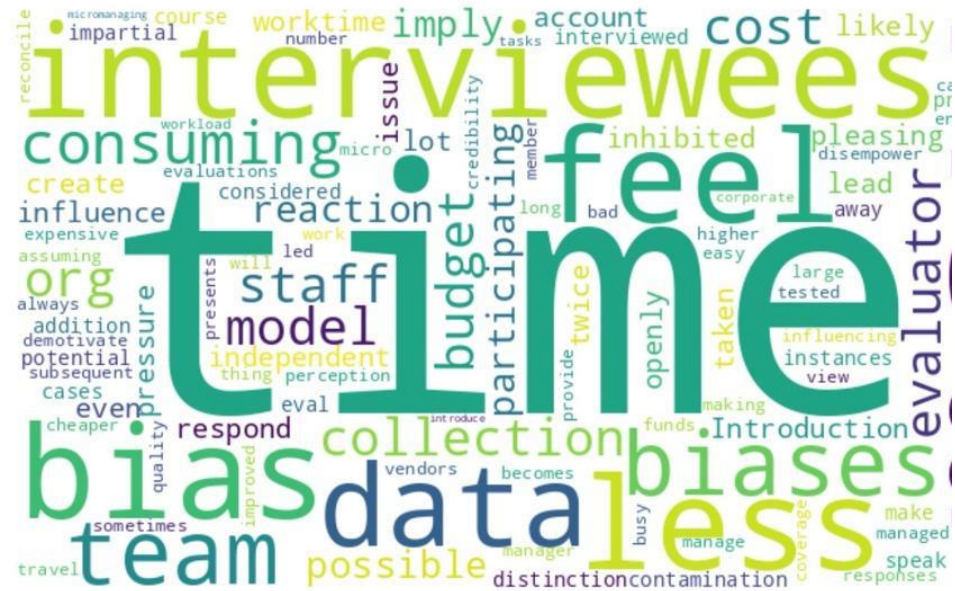
Even if many respondents reported a favorable outlook on direct participation, careful consideration is required to balance these benefits with potential downsides. Many respondents expressed concerns that participation could introduce **bias**, as managers' association with the organization might inadvertently influence respondents' answers. **Over-involvement or micromanagement** were also perceived as

potential demotivators for consultants, possibly creating tension, reducing their autonomy, or inhibiting performance.

The most mentioned drawbacks also included the significant **time and budget** demands associated with managers' participation, which could strain resources and add to their workload.

Figure 23. Word cloud on the pros for participating in data collection

Figure 24. Word cloud on the 'cons' for participating in data collection



Another open-ended question explored respondents' perspective on the **main challenges of data collection** for evaluation. Figure 25 presents a word cloud based on the responses. It is immediately evident that **time** is the most frequently mentioned challenge. Participants reported that time pressures often lead to insufficient data or rushed analyses. These constraints are frequently mentioned alongside **budget** constraints, as financial limitations on travel, hiring qualified team members, and allocating sufficient field time significantly hinder comprehensive data collection and negatively affects **data quality**.

Another prominent theme emerging from the answers and clearly reflected in the word cloud concerns several challenges related to **access**. Respondents highlighted difficulties in reaching hard-to-access areas such as rural regions, conflict zones, or areas requiring significant travel. Access issues were also reported in engaging vulnerable or underrepresented groups, such as indigenous people, migrants, or marginalized communities. Additionally, respondents noted challenges in securing interviews with key stakeholders, particularly those outside immediate organizational networks or in sensitive sectors. Finally, delays in obtaining necessary documents or background information from evaluands were also frequently mentioned as significant obstacles. Another access challenges relates to possible language barriers.

Survey participants also expressed concerns with implementation of **adequate methodology**, as they often encounter challenges in ensuring appropriate sampling strategies to capture diverse perspectives and avoid respondent fatigue or bias, as well as difficulty ensuring that qualitative and quantitative data are effectively integrated while maintaining credibility and transparency in findings.

Some survey participants highlighted challenges in data collection related to **bias and reliability**. These include the risk of response bias, communication barriers, inconsistency in responses caused by language, cultural nuances, or lack of accurate recall from respondents. Additionally, some respondents also mentioned that the **limitations on free speech** and a reluctance to challenge donors or the government often undermine the impartiality, depth and objectivity of the data collection efforts.

Figure 25. Word cloud for the three main challenges in data collection

2.6 Evaluation Reports

This section of the survey focused on **evaluation managers' involvement in final evaluation reports**, as well as the frequency and satisfaction levels regarding internal and external peer reviews. About half of the respondents reported contributing to final evaluation reports, while the other half indicated they rarely or never contribute. Managers from implementing organizations were the most likely to contribute, followed by managers from UN agencies and international research organizations, who often act as contributors. Most respondents from government entities and all of those from donor organizations and other multilateral organizations and funds reported rarely or never directly contributing to final evaluation reports (Figure 26).

With regards to the sections of the report that the respondents contribute to, Figure 27 shows that respondents tend to contribute equally across all parts of the report. These include the background/context, the evaluation methodology, results and key findings, as well as recommendations and conclusions. Additional sections mentioned in the comments include the executive summary and annexes.

Approximately half of the respondents reported having sufficient time to properly review evaluation deliverables, while about 40% indicated that they do not have enough time and 10% remained neutral. These results were consistent across most organization types, except for donor organizations, where respondents were more likely to report a lack of adequate time (Figure 28).

Figure 26. Do you contribute to the original writing of the report? (N-43)

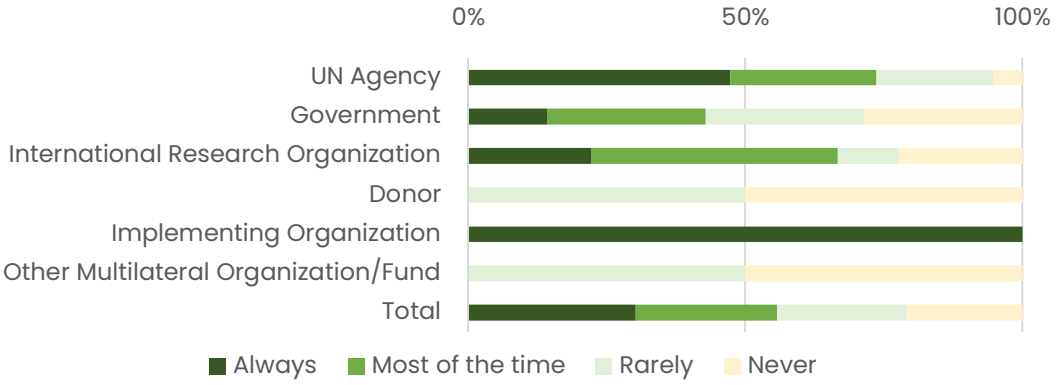


Figure 27. Which parts do you contribute to? (N-34)

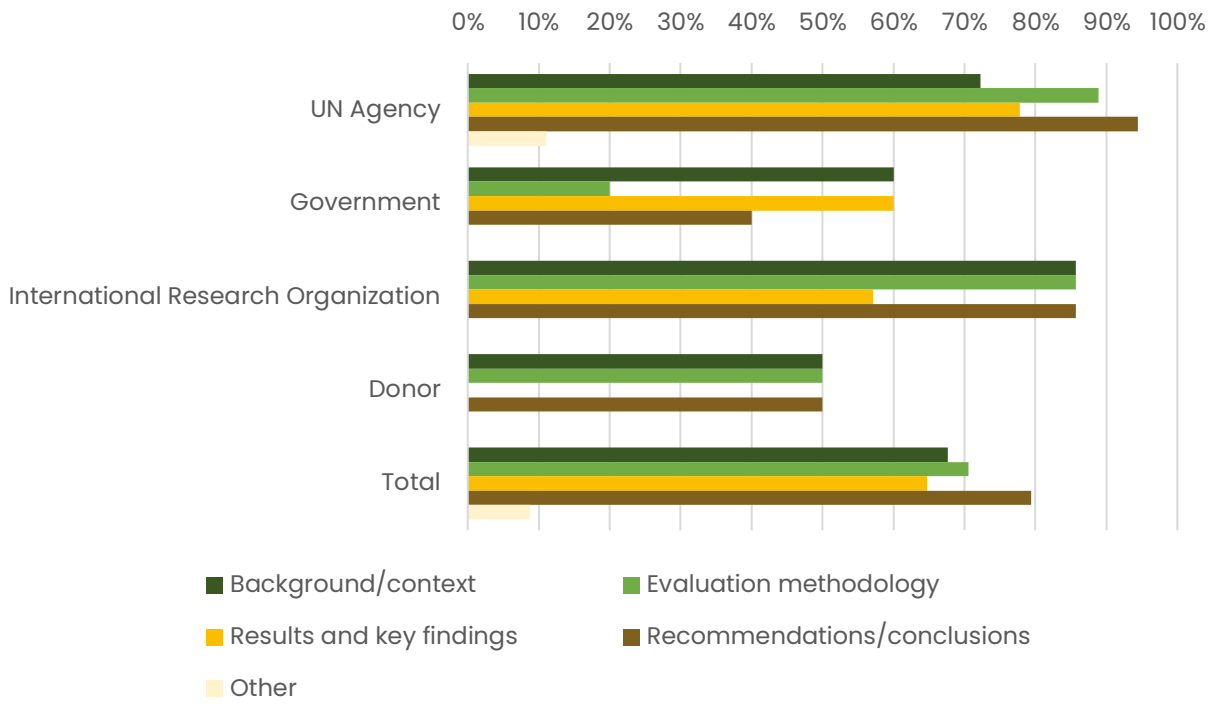


***Figure 28. Do you agree with the statement*: "As Evaluation Manager, I usually have enough time to properly review the evaluation deliverables (reports, sub-studies, analysis...)"? (N-42)**

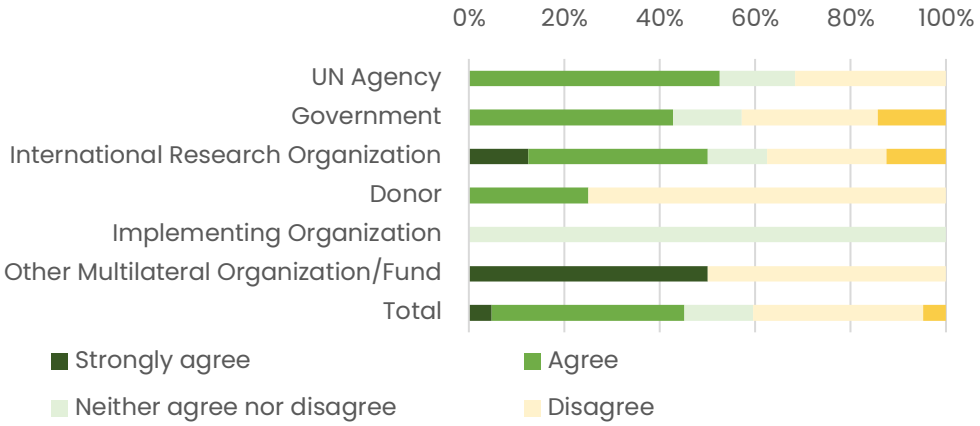


Figure 28 reveals a strong involvement of ***internal* peer reviewers**, with nearly 90% of respondents consistently submitting draft evaluation reports to them. According to respondents, this is especially prevalent in UN agencies, international research organizations, implementing organizations, and other multilateral organizations and funds. This figure decreases to about 70% among respondents from government entities and to 50% for those from donor organizations. Figure 30 shows that nearly all respondents regard the contribution of internal peer reviewers as a significant added value to the evaluation report.

Reliance on ***external* peer reviewers** is less common compared to internal reviewers, with approximately 50% of respondents reporting that they consistently submit draft evaluations to external peer reviewers. A clear distinction in practices emerges among the organizations: more than half of survey participants from UN agencies, government entities and international research organizations rely on external peer reviewers, whereas respondents from donor organizations, implementing organizations and other multilateral organizations and funds rarely or never do so (Figure 31). Nonetheless, as shown in Figure 32, almost all respondents generally consider the contribution of external peer reviewers to be highly valuable, even if such involvement is not common in their organizations.

Figure 29. Do you submit the draft evaluation report to internal peer reviews for feedback? (N-42)

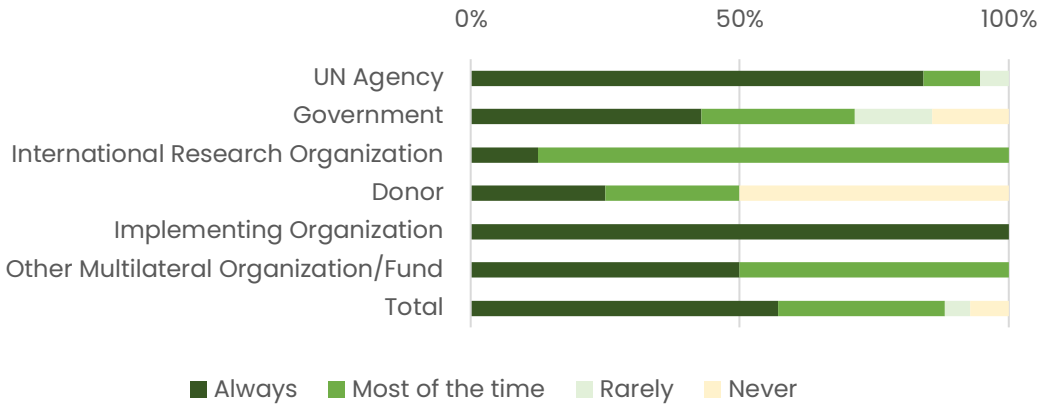


Figure 30. Do you agree with the statement: "The contribution of internal peer reviewers is an added value to the evaluation report"? (N-39)

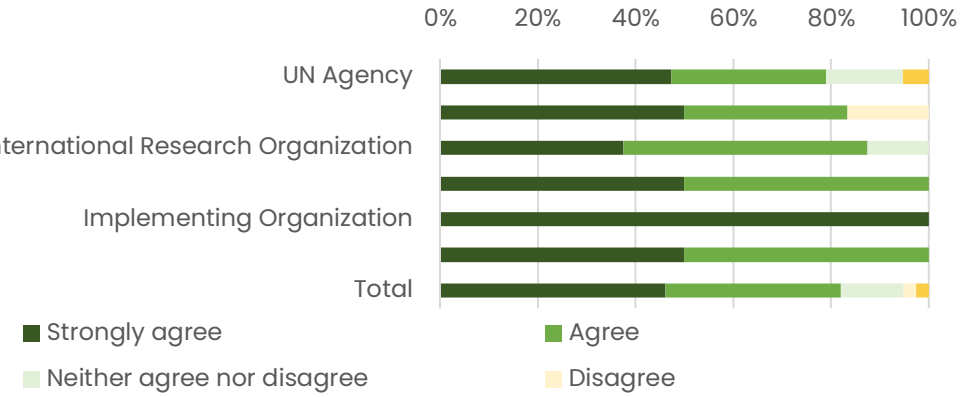


Figure 31. Do you submit the draft evaluation report to external peer reviewers for feedback? (N-43)

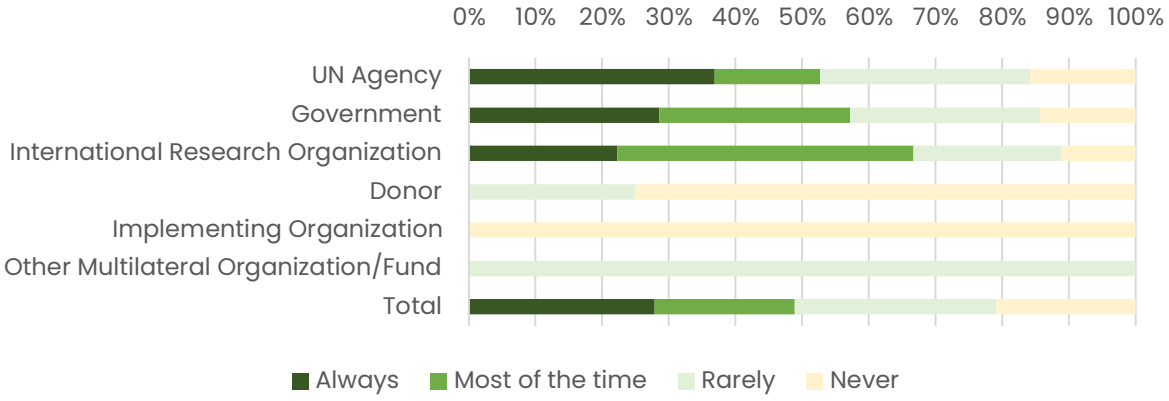
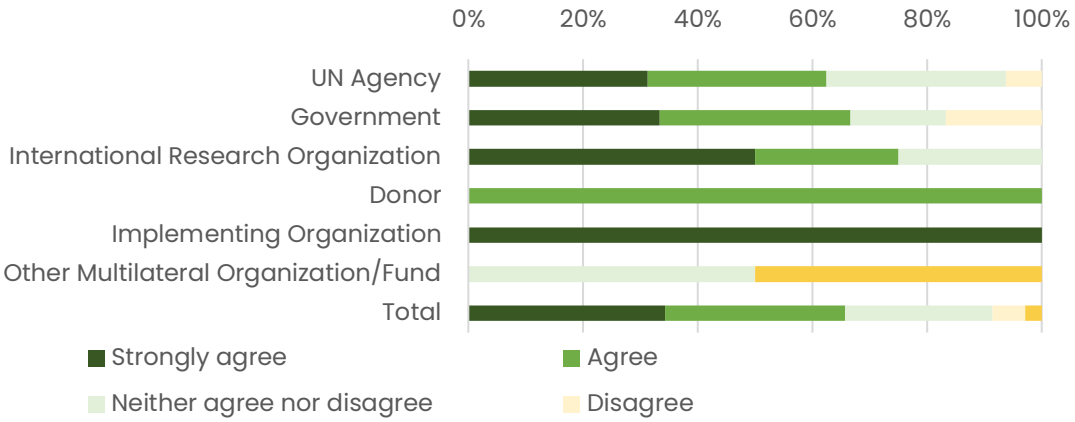


Figure 32. Do you agree with the statement "The contribution of external peer reviewers is an added value to the evaluation report"? (N-35)



Two survey questions explored the use of **Artificial Intelligence (AI)** in evaluation. Just over half of respondents indicated that AI is either used for selective tasks or by external consultants, while the other half reported that it is not used for evaluations (Figure 33). Respondents from UN agencies were more likely to use AI directly for selective tasks, whereas those from donor organizations, government entities, and other multilateral organizations and funds reported that AI is primarily used by external consultants. International research organizations were the least likely to use AI.

Figure 34 reveals that when AI is employed for tasks such as note-taking and summarizing, a quality check is performed by about half of the respondents, particularly within UN agencies. In contrast, quality checks are less common among other types of organizations included in the survey.

Figure 33. To what extent is AI used in evaluations you are involved in? (N-43)

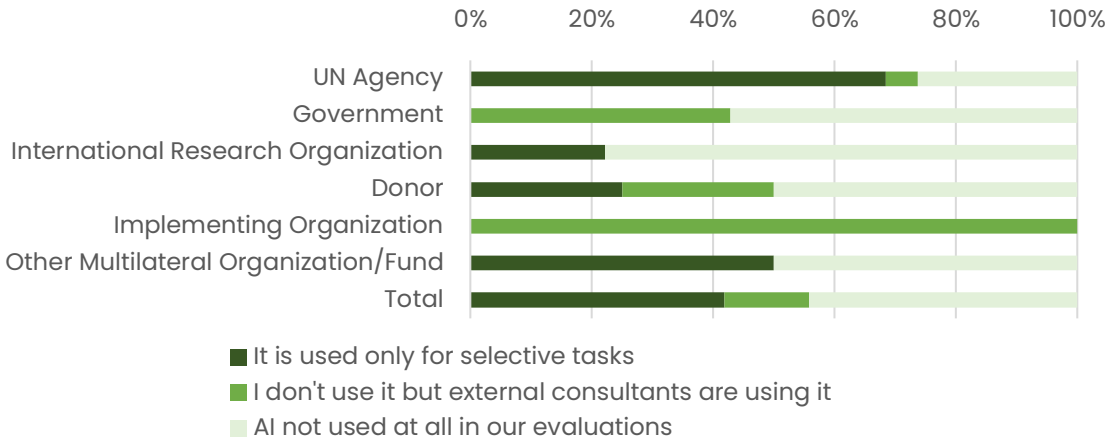
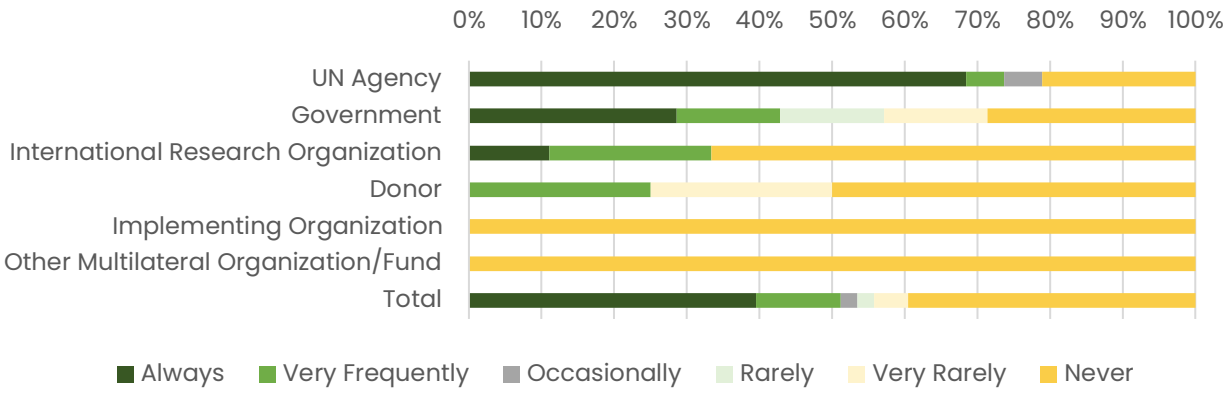


Figure 34. If AI is used for notetaking and summarizing, is there a quality check performed after the notes are produced? (N-43)



2.7 Publication and Use of Evaluation Reports

This section of the survey focused on the publishing (in the public domain) and dissemination of evaluation reports. It examines the time required to publish reports, the criteria for publication, who is responsible for presenting results to stakeholders, and how respondents rate the use of evaluative evidence within their organizations.

According to survey participants, evaluation reports are always or almost always published in approximately 80% of the cases (Figure 35). This is especially true in UN agencies and implementing organizations. Respondents from government entities, international research organizations, donors and other multilateral organizations and funds do not show a clear trend in whether reports are published.

Nearly 60% of respondents reported a period of less than three months between the validation of the report and the publication, about 30% reported between six and 12 months and a few reported that it takes more than a year (Figure 36). Shorter time frames are more common among respondents from UN agencies, international research organizations, followed by those from government entities and other multilateral organizations and funds. Implementing organizations respondents reported to have slightly longer time frames.

Approximately a fifth (22%) of respondents reported that the decision to publish an evaluation report is made before the evaluation begins. Another 12% indicated that the decision is taken after the evaluation, while about 35% stated that it depends on the type of evaluation, and 30% cited other criteria (Figure 37). Many respondents noted that publishing all reports is part of their organization's policy, emphasizing principles of transparency, accountability, and knowledge -sharing. However, some respondents highlighted that reports are not published if they are deemed confidential or contain sensitive information.

According to survey participants, evaluation results are primarily presented to governing bodies and donors by the evaluation manager and/or the evaluation team, either jointly or by one of the two individually (Figure 38). In UN agencies, evaluation managers are predominantly responsible for this task, whereas in donor organizations, it is typically handled by the evaluation team. Responses from other organizations indicate a mix of approaches, with no clear preference for a specific method.

Respondents were asked to rate the use of evaluative evidence for decision-making processes, such as planning and mid-course corrections, within their organizations. The overall rating was relatively unsatisfactory, with an average score of 3.3 (Figure 39). Respondents from donor organizations provided the highest average score (3.8), followed by UN agencies (3.5). International research organizations and other multilateral organizations and funds both averaged a score of 3.0, while government entities reported the lowest score (2.8). These scores were further contextualized in the comments section, where respondents highlighted that the utility of evaluations depends on factors such as the type of evaluation and its timing—being higher for mid-term reviews (MTRs) compared to final evaluations. Some respondents also noted difficulties in making overall judgments due to a trade-off between quantity and quality, indicating that an excessive number of evaluations can reduce their overall impact and recommendation uptake.

Figure 35. Are your evaluation reports published? (N-42)

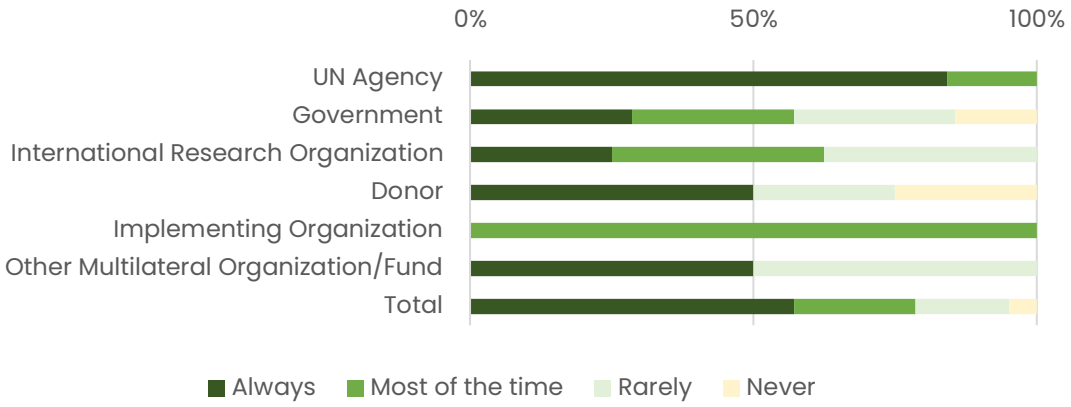


Figure 36. How long does it take from validation of the report to its publication? (N-39)

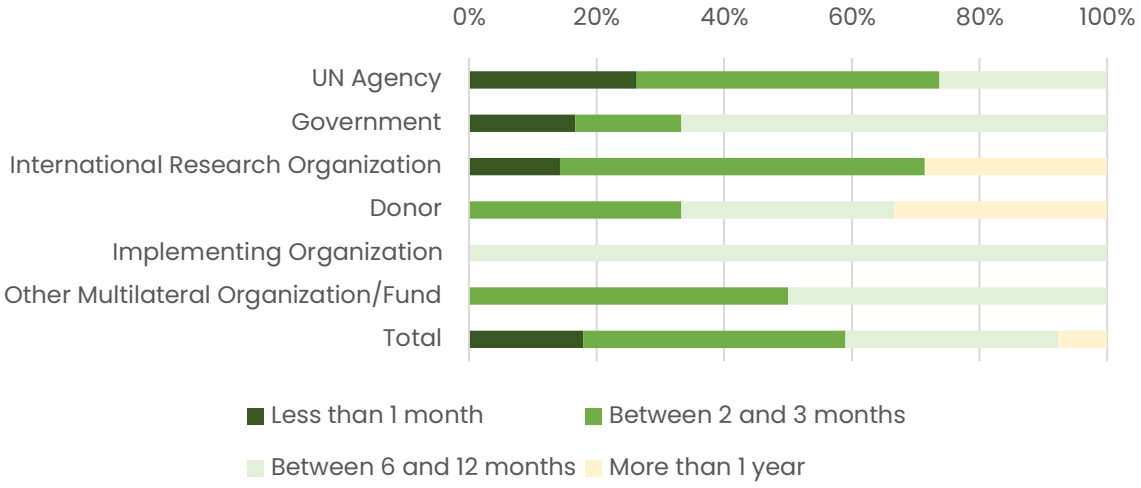


Figure 37. The criteria for publishing the evaluation report is... (N-40)

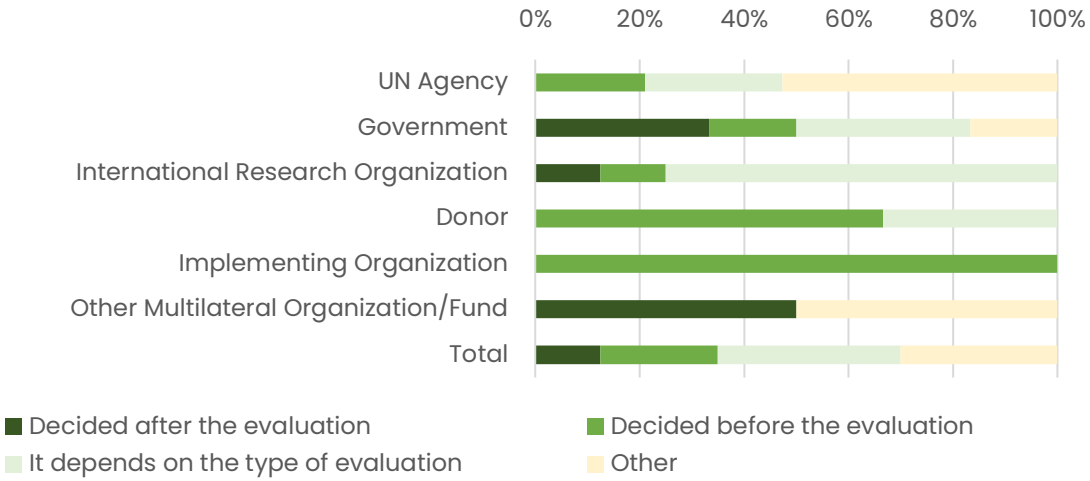


Figure 38. Who presents the evaluation results to governing bodies and/or donors (N-36)?

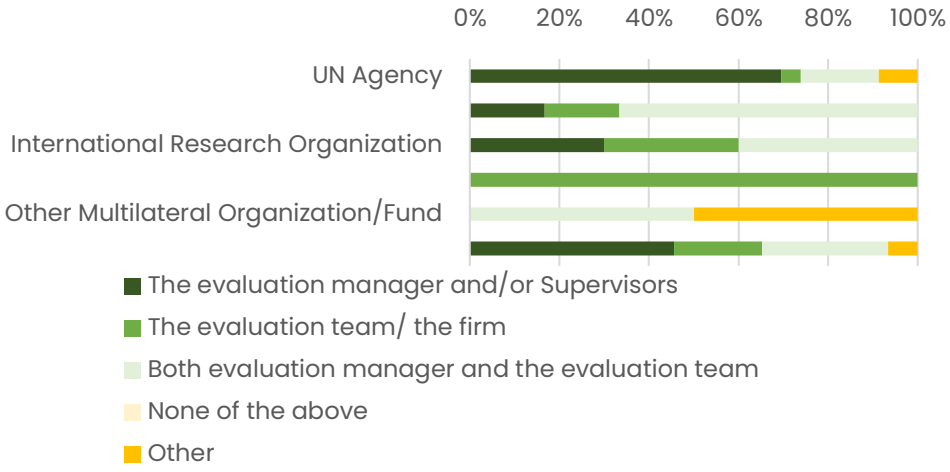


Figure 39. In your organization, how would you rate the use of evaluative evidence for decision making (planning, mid-course correction...) (N-37)

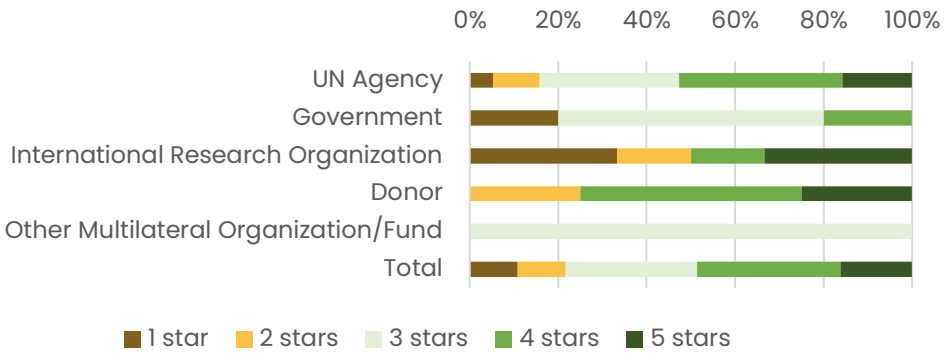
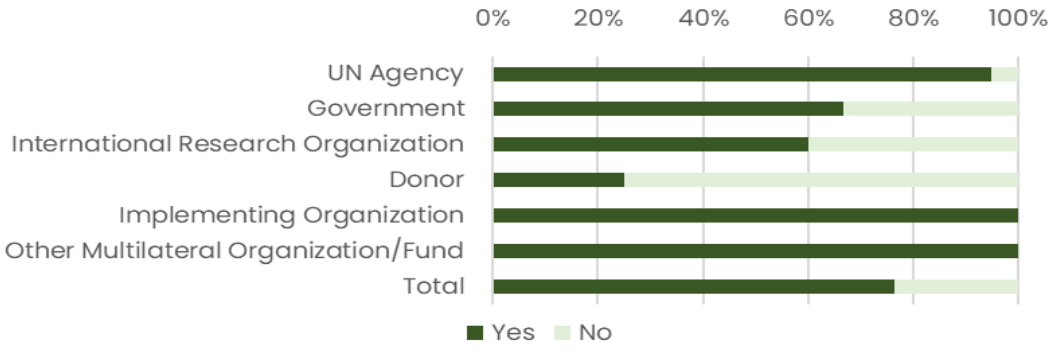


Figure 40 illustrates that the responsibility for tracking recommendations frequently falls within the evaluation office, entity or service, as reported by over 60% of respondents. In UN agencies and international research organizations, some respondents indicated that this responsibility can also be handled by management, other organizational divisions, or evaluands.

Figure 40. Is the MR developed for all evaluations? (N-38)



2.8 Management Response and Tracking

A final part of the survey examined practices related to the MR. 76% of respondents reported that a MR is developed for all evaluations, it is a standard practice in UN agencies, implementing organizations and other multilateral organizations and funds. About 60% of respondents from government entities and international research organizations reported regularly implementing this practice, compared to only 25% of survey respondents from donor organizations (Figure 41).

Just under half of the respondents indicate that the MR usually takes less than one month, while about 38% reported that it takes more than two months. A few respondents, particularly from implementing organizations, donor organizations and international research organizations were unsure about the duration. The process appears to be a quicker process, often under one month, in government entities, UN agencies and other multilateral organizations and funds (Figure 42).

Figure 43 shows that the MR is reported to be published either always or most of the time by just over half of the respondents, while 20% reported it is rarely or never published. Systematic publication is most frequent in UN agencies, where 90% of respondents reported it is a standard practice. This is followed by other multilateral organizations and funds, with 50% reporting systematic publication. Around 30% of respondents from government and international research organizations reported it as a regular practice, while it appears to be a rare occurrence in donor organizations.

The MR is reported to be published at similar rates either in the same document of the evaluation report (31%), or in a separate document at the time of report publication (38%) or in a separate document later (34%). The second option is more common in international research organizations and UN agencies, while government entities are more likely to publish it later (Figure 44).[20] Respondents from UN agencies noted that MRs are often published on the organization's website. One respondent further highlighted that there is an institutional dashboard that collects all recommendations from independent evaluations alongside their MRs. This system is also used to track the implementation of recommendations.

70% of survey participants indicated that their organization has a system for tracking the implementation status of the MR (Figure 45). The system is most common in UN agencies (94%), followed by international research organizations (57%), government entities (50%), other multilateral organizations and funds (50%) and donor organizations (25%). However, only 30% of respondents reported that the MR tracking system is publicly accessible (Figure 46).[21] Notably, all respondents from government entities reported that it is never publicly available.

[20] The total includes responses from development banks, implementing organizations and other multilateral organizations and funds. They were not presented separately as one answer was received for each of these three categories.

[21] The total includes responses from development banks, donor organizations, implementing organizations and other multilateral organizations and funds. They were not presented separately as one answer was received for each of these four categories.

Figure 41. Is the MR developed for all evaluations? (N-38)

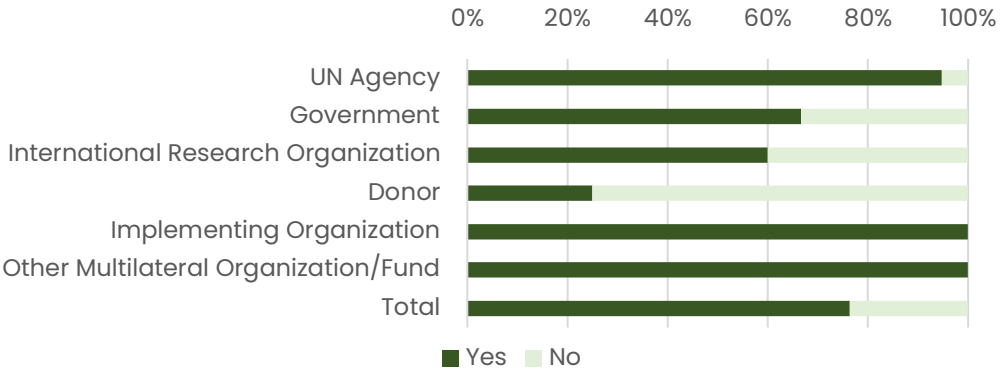


Figure 42. How long does the development of MR usually take? (N-40)

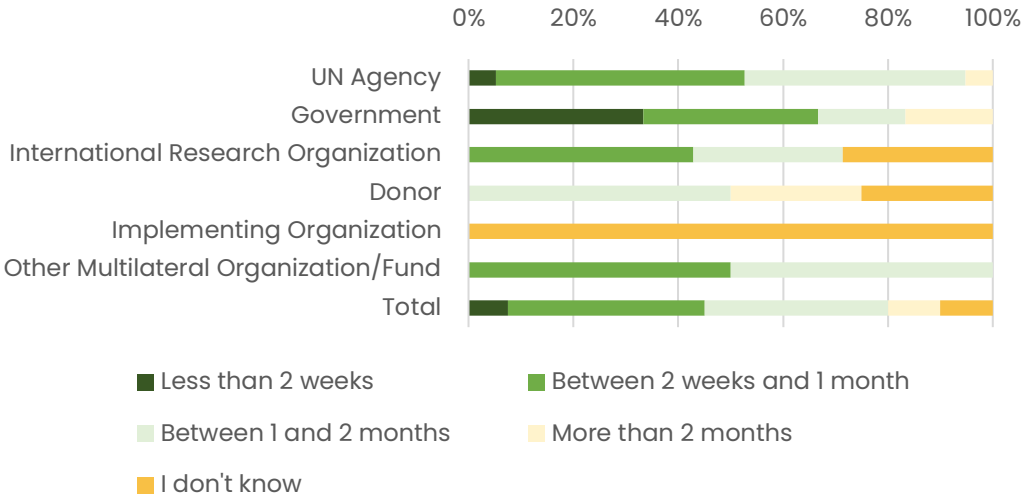


Figure 43. Does the MR usually get published? (N-40)

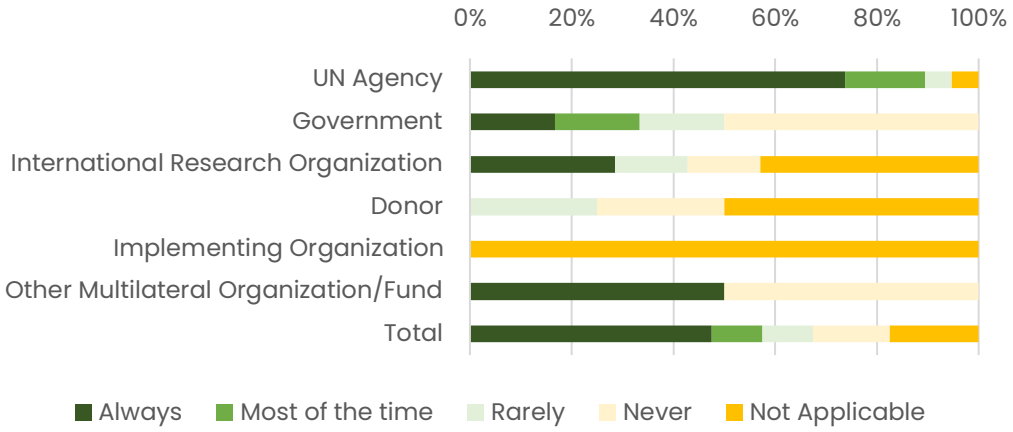


Figure 44. If yes, where does the MR get published (N-32)

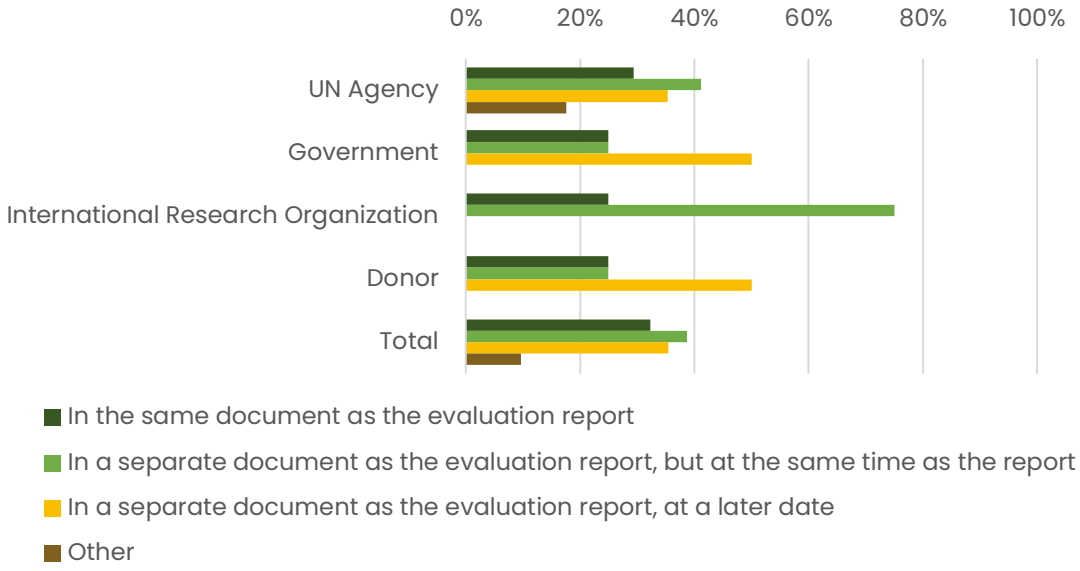


Figure 45. Does your organization have a system for tracking status of implementing MR? (N-40)

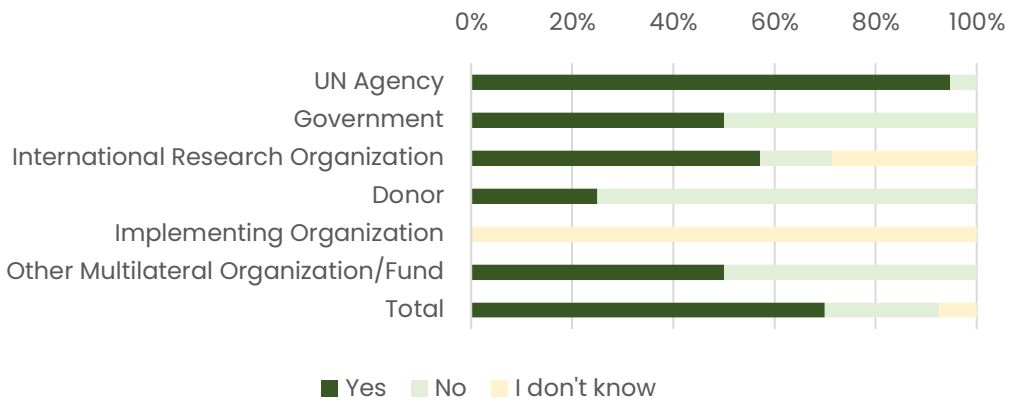


Figure 46. Is the MR tracking system publicly available/accessible? (N-27)

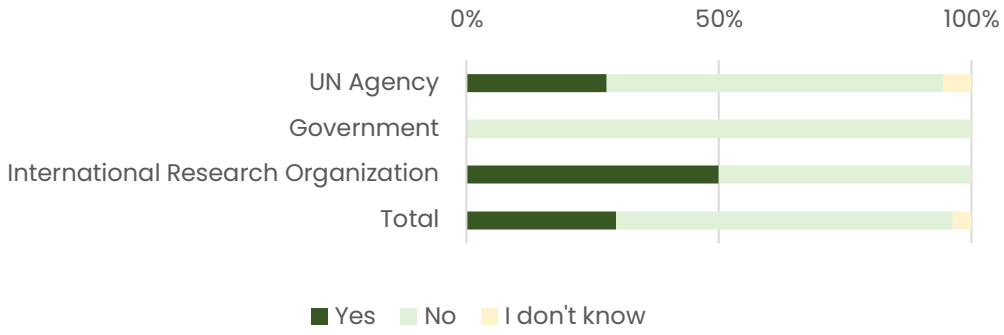
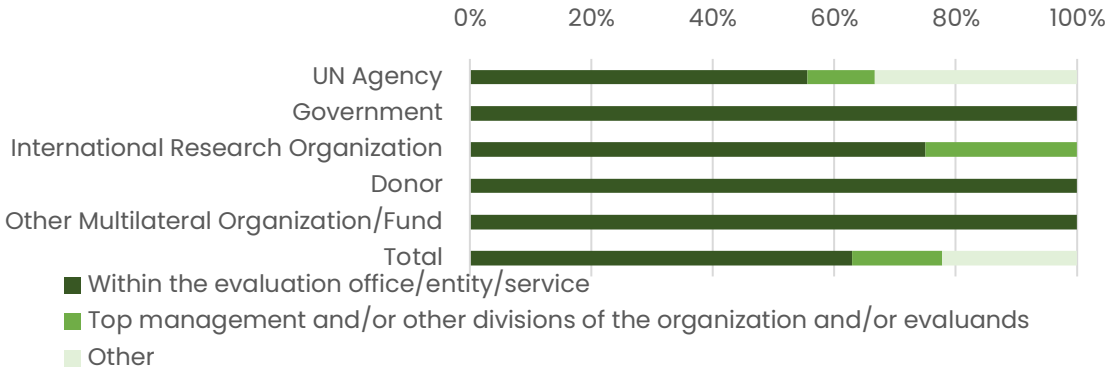


Figure 47. If yes, who oversees the tracking of recommendations? (N-27)



3 Results of the Evaluations Mapping Across Peer Organizations

This section provides an overview of how evaluations have been conducted across several peer organizations, based on a review of 100 sampled evaluation reports. The analysis focuses on three aspects: duration of evaluations, country coverage, and team composition. Where possible, data is disaggregated by evaluation type, below is a summary and mapping details and data can be found in Annex 4.

3.1 Duration of Evaluations

Out of the 100 evaluations reviewed, information on the approximate duration was available for 78. Among these, the average duration was approximately 11 months, suggesting that most evaluations are year-long processes. However, there was significant variation, with durations ranging from as short as one month to as long as two years and two months. These variations reflect differences in scope, complexity, and resources available for each evaluation.

3.2 Country Coverage

Given the disruptions caused by the COVID-19 pandemic, many evaluations could not include country visits. To capture intended reach, the number of countries visited was supplemented with the number of country case studies included. This broader definition allowed for a representation exploration of the geographic scope of evaluations during the pandemic years. Data on country coverage was available for 56 evaluations. On average, each evaluation covered 3.7 countries. Most reports focused on a single country, but some had a much wider scope—up to 43 countries.

3.3 Team Composition

Information on the number of team members involved in each evaluation was available for 93 of the 100 reports. Team sizes ranged widely—from a single evaluator to teams of up to 42 people. The average team size was eight members, reflecting a mix of smaller and larger evaluation teams depending on the scale and requirements of the evaluation.

4 Conclusions and Recommendations

The management of independent evaluations significantly influences the utilization of evaluation results. The study primarily based on online survey results about evaluation management styles examined how various international organizations manage independent evaluations, identifying different practices employed across the evaluation process and the key challenges encountered. Although no direct correlation was established between specific management models and respondents' perceptions of evaluation use (based on 66 valid responses), insights from the literature review, the EvalforEarth discussion, and the evaluation mapping enable us to draw the following conclusions, structured according to the typical phases of an evaluation.

Evaluation Terms of Reference

The survey results indicate that evaluation management constitutes a significant portion of respondents' duties. Except for government institutions, evaluation managers are typically responsible for ToRs. In most cases, ToR development takes fewer than ten days. Additionally, evaluation managers commonly conduct and/or oversee the design of the evaluation approach, methodology, and the formulation of key questions. A participatory approach is generally followed in developing ToRs, with the evaluand team cited as the most involved actor across all organizations—except among donors, who primarily identify the commissioning entity as the key additional stakeholder in the process.

EAs remain limited in practice. Only 35% of respondents reported conducting such assessments consistently or most of the time, while the majority noted that these occur sporadically (See EA Portal of CGIAR). According to UNEG Standard 4.2, an EA should be undertaken as an initial step to enhance the likelihood that an evaluation will yield timely and credible information for decision-making.

The literature on evaluation use in agricultural research emphasizes that evaluations must be intentionally designed to align objectives, stakeholders, and the anticipated applications of results. Tools such as ToRs play a key role in defining the purpose and intended use of evaluations. However, persistent challenges in systematizing and managing evaluation results highlight the need for structured processes to document feedback and ensure continuous utilization.

UNEG Standard 4.6 (UNEG, 2016) underscores the importance of Inclusive and diverse stakeholder engagement in the planning, design, conduct and follow-up of evaluations to ensure ownership, relevance, credibility and the use of evaluation. Reference groups and other stakeholder engagement mechanisms should be designed for this purpose.

Findings and Contracting the Evaluators

UN agencies have a strong preference for individual consultants, while the other organizations highlight that it depends on the type of evaluation, evaluand and context. Individual consultants have the advantages of an easier contracting process and lower costs. Individual consultants are often preferred for project level evaluation, which typically have lower budgets. Firms, on the other hand, can be deemed to provide QA, backstopping, credibility and more suited for complex or large-scale evaluations.

Finding and hiring the right team is reported as difficult by survey respondents. Challenges include time constraints, low availability of SMEs, budget constraints, and long bureaucratic processes. Respondents noted the difficulty of finding consultants with both subject matter expertise and evaluation knowledge, combined with strong analytical and writing skills. Another significant challenge relates to managing

individuals, which can be time-consuming. Coupled with challenging team dynamics, this can negatively affect the quality of the evaluation.

Data collection and inquiry

A common practice across organizations is the involvement of evaluation managers in the design of data collection. However, participation in fieldwork varies. Donor organizations, implementing organizations, and other multilateral organizations or funds rarely engage in field missions. In contrast, UN agencies, government entities, and international research organizations participate more frequently. Respondents' involvement includes conducting interviews, facilitating focus groups, and engaging in other data collection activities.

Survey results (open-ended questions) and the EvalforEarth discussion highlight the importance of calibrating the role of evaluation managers. Specifically, regarding data collection, survey results indicate that many evaluation managers see direct participation as a means of improve the overall quality of evaluations. However, concerns about bias and micromanagement are also widely expressed. The most mentioned drawbacks include the significant time and budget associated with managers' participation, which can strain resources and add to their workload.

The EvalforEarth discussion underscored the necessity of preserving the independence of evaluation teams. Due to potential biases and conflicts of interest, respondents suggest that managers adopt a facilitative and supportive role rather than direct involvement in data collection. This approach enhances the evaluation process, particularly when evaluators require extensive background information. The inception phase is key to establishing a strong foundation for a successful evaluation, ensuring proper orientation and clearly defining the roles of evaluation managers.

Survey findings reveal several interconnected challenges in the data collection phase. Time constraints frequently lead to compromised data quality and rushed analyses. Budget limitations affect travel, the hiring of qualified personnel, and the allocation of sufficient field time, with evaluation mapping exercise indicating that the average duration for most sampled evaluations (78 out of 100) was just under a year (11 months).

Access difficulties include reaching remote or conflict-affected areas, engaging marginalized groups, and securing interviews with key stakeholders outside established networks. Methodological concerns revolve around ensuring diverse sampling strategies, integrating qualitative and quantitative data effectively, and addressing respondent fatigue and bias. Finally, bias and reliability issues arise due to risks such as response bias, communication barriers, and cultural nuances, all of which can impact data impartiality and objectivity.

Reporting and Communication of Results

Internal peer reviews are viewed by survey respondents as a valuable means to improve the quality of evaluation reports. External peer reviews are also considered highly valuable, though this practice is not widespread. Regarding the use of AI, respondents from UN agencies are more likely to use AI directly for specific tasks, while those from donor organizations, government entities, and other multilateral organizations and funds report that AI is primarily used by external consultants. International research organizations are the least likely to use AI. AI is commonly employed for tasks such as note-taking and summarizing, and a quality check is performed by about half of the respondents, particularly within UN agencies. In contrast, quality checks are less common among other types of organizations included in the survey.

Regarding the later stages of independent evaluations, all organizations publish their reports. However, there is significant variation in the perceived effective use of these reports across organizations, with less consistent use, especially within government entities and donor organizations. The number of reports produced, and the time required for their preparation have proven to be bottlenecks that hinder the effective use of results. A good practice, particularly in the management of agricultural research, is to ensure that reports and recommendations are accessible to the entire organization, except in cases where the information is sensitive.

The EvalforEarth discussion highlights the value of effective communication, such as initial and final meetings between evaluation managers and teams, to provide contextual insights and refine conclusions.

Use, Management Response and Tracking

Respondents' overall satisfaction with the use of evaluative evidence in decision-making processes within their organizations was low, with government entities reporting the lowest scores. Respondents highlighted that the utility of evaluations depends on factors such as the type of evaluation and its timing—being higher for mid-term reviews compared to final evaluations. A high number of evaluations can reduce their overall impact and the uptake of recommendations.

The UNEG, in its Norms and Standards, state that organizations should ensure that appropriate mechanisms are in place to ensure management response to evaluation recommendations, and that these mechanisms outline concrete actions to be undertaken in the MR and in the follow-up to recommendation implementation.

A benchmarking review conducted by IAES Evaluation Function (IAES. 2025) mapped existing MR structures, processes, and review methodologies, exploring best practices in implementation, oversight mechanisms, and tracking systems to support evaluation uptake. The review found that common challenges in tracking actions to recommendations are mainly: resource constraints, operational complexities, cultural resistance to change, data limitations, and gaps in tracking and strategic alignment. The review underscores the importance of having well-defined systems and processes for managing responses to evaluation recommendations and the need for stronger integration of evaluation findings and recommendations into strategic planning and decision-making processes. Additionally, robust knowledge management and dissemination mechanisms are key for ensuring that lessons learned from evaluations are shared and applied across the organization and beyond.

Literature identified a clear gap in the management and systematization of evaluation results. The growing global emphasis on RRI and RRA and responsible investment, and mission-oriented research, combined with the SDGs, reinforces the importance of evaluation as a tool for social transformation (Von Schomberg, 2019). Establishing an organizational culture focused on societal impact means incorporating evaluation not just as a bureaucratic requirement, but as a catalyst for change. Pinto & Bin (2024) highlight that agricultural R&D&I institutions can use evaluation results to recalibrate research focus, optimize project design, and influence resource allocation, supporting socially beneficial innovations and promoting positive impacts across multiple dimensions.

The use of evaluative evidence is primarily a management matter. To foster use, effective management processes should include highly participatory approaches, ensuring that evaluation design, objectives, and scope are tailored to the specific context and available resources. Evaluation managers play a key role in shaping the evaluation methodological framework from the outset and should be well-trained and equipped for this task. Conducting an EA can help save time and manage expectations. Balancing independence and evaluation quality requires a carefully designed and clearly communicated distribution

of roles, such arrangement should be adapted to the specific evaluation and context. Mid-term evaluations are more likely to drive course corrections. Additionally, tracking systems should be accessible and effectively used to monitor progress and inform decision-making.

The Evaluation Function of CGIAR will present the results of the study in different evaluation conferences/fora to trigger a discussion among agencies about the topics of managing evaluations and collect more insights. Internally, the overall aim is for CGIAR to adapt and tailor its evaluation practices towards more use of independent evaluation advice and inform the future revision of the Evaluation Framework and Policy.

Figure 48. Recommendations for effective evaluation management

A.	**Evaluation Design and Development of a ToR**
1.	Provide continuous training for evaluation managers to strengthen their capacity in designing evaluation approaches and methodologies.
2.	Conduct evaluability assessments (EAs) to clarify what can and cannot be evaluated, manage expectations, save resources, and enhance feasibility.
3.	Use reference groups and other stakeholder engagement mechanisms to strengthen credibility and ownership.
B.	**Findings and Contracting the Evaluators**
4.	Develop a network database of qualified consultants to reduce hiring difficulties and streamline processes.
C.	**Data Collection and Inquiry**
5.	Clearly define roles and responsibilities, particularly for managers and stakeholders. Clear communication and collaborative planning between evaluation managers and external evaluators is key to the process.
D.	**Reporting and Communication of Results**
6.	Ensure that evaluation reports and recommendations are accessible for different audiences to enhance their use.
7.	Foster effective communication between evaluation managers and teams through regular meetings to provide contextual insights and refine conclusions and recommendations.
E.	**Use, Management Response and Tracking**
8.	Conduct mid-term reviews, as they are more likely to be used by stakeholders to perform corrective actions.
9.	Tracking systems for management responses should be publicly accessible to promote transparency and accountability, particularly in a context of responsibility-driven research and development (R&D).
10.	Strengthen mechanisms to ensure timely follow up on recommendations, integrating them into strategic planning and decision making.

References

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. American Journal of Evaluation, 37(4), 568–579. https://doi.org/10.1177/1098214016665164.

Better Evaluation. (2023). Evaluating capacity development. https://www.betterevaluation.org/methods-approaches/themes/evaluating-capacity-development.

Chelimsky, E. (1977). Improving the Use of Evaluation: An Agency Perspective. The Prison Journal, 57(1), 13–18. https://doi.org/10.1177/003288557705700103.

Chelimsky, E. (2015). A strategy for improving the use of evaluation findings in policy. Evaluation use and decision making in society: A tribute to Marvin C. Alkin, 73–89.

CGIAR Independent Advisory and Evaluation Service (IAES). (2025). Management Response Systems to Evaluations: Benchmarking Review. Rome: IAES Evaluation Function.

CGIAR. (2023). Terms of Reference of CGIAR's Independent Advisory and Evaluation Service IAES. https://storage.googleapis.com/cgiarorg/2022/08/TOR-IAES-Approved-4Oct2018.pdf.

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the Capacity to Do and Use Evaluation. New Directions for Evaluation, 2014(141), 7–23. https://doi.org/10.1002/ev.20076.

Evaluation Cooperation Group (2012). Big Book on Evaluation Good Practice Standards. https://www.ecgnet.org/document/ecg-big-book-good-practice-standards.

Ferré, M., Blundo-Canto, G., Rodrigues, G. S., Ramírez-Gómez, M.-M., Vedovoto, G. L., Agudelo-Chocontá, B.-E., Marques, D. V., Flores, R. M. V., Rodríguez-Borray, G.-A., & De Souza, M. O. (2025). Culture of impact in agricultural research organisations: What for and how? Research Policy, 54(1), 105140.

Ferre, M., Gomez, M. M. R. G., Vedovoto, G. L., Canto, G. B., Chocanta, B. E. A., Marques, D. V., Rodríguez-Borray, G.-A., Souza, M. O. de, Urriago, A. R. V., Rodrigues, G. S., Sánchez-Lozano, J.-I., Zambrano-Moreno, G.-S., Arias, M. a. L., Zapata-Tamayo, C.-A., Goulet, F., Proietti, C., & Romémont, A.de. (2023). Culture of impact in agricultural research organizations: What for and how? - insights from a cross-analysis of Cirad, Embrapa, and Agrosavia institutes. https://www.alice.cnptia.embrapa.br/alice/handle/doc/1159885.

Hall, A., Sulaiman, V. R., Clark, N., & Yoganand, B. (2003). From measuring impact to learning institutional lessons: An innovation systems perspective on improving the management of international agricultural research. Agricultural systems, 78(2), 213–241.

Horton, D., & Mackay, R. (2003). Using evaluation to enhance institutional learning and change: Recent experiences with agricultural research and development. Agricultural Systems, 78(2), 127–142.

Joly, P.-B., Colinet, L., Gaunand, A., Lemarié, S., & Matt, M. (2016). Agricultural research impact assessment: Issues, methods and challenges.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. American Journal of Evaluation, 33(3), Artigo 3.

Mackay, R., & Horton, D. (2003). Expanding the Use of Impact Assessment and Evaluation in Agricultural Research and Development. Agricultural Systems, 78, 143–165. https://doi.org/10.1016/S0308-521X(03)00124-0

Mertens, D. M. (2008). Transformative research and evaluation. Guilford Press.

Milzow, K., Reinhardt, A., Söderberg, S., & Zinöcker, K. (2019). Understanding the use and usability of research evaluation studies. research evaluation, 28(1), 94–107.

Morgan Jones, M., Grant, J., & RAND, E. (2013). Making the grade: Methodologies for assessing and evidencing research impact. Dean A, Wykes M, Stevens H, editors, 7.

Morgan Jones, M., Manville, C., & Chataway, J. (2017). Learning from the UK's research impact assessment exercise: A case study of a retrospective impact assessment exercise and questions for the future. The Journal of Technology Transfer, 1–25.

OECD. (2019). Evaluation Criteria. https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

OECD. (2011). Busan Partnership for Effective Development Co-operation: Fourth High Level Forum on Aid Effectiveness, Busan, Republic of Korea, 29 November - 1 December 2011, OECD Publishing, Paris, https://doi.org/10.1787/54de7baa-en.

Parker, C., Scott, S., & Geddes, A. (2019). Snowball Sampling. SAGE Research Methods Foundations. http://methods.sagepub.com/foundations/snowball-sampling.

Patton, M. Q., & Horton, D. (2009). Utilization-focused evaluation for agricultural innovation. ILAC Brief.

Patton, M. Q. (2008). Utilization-focused evaluation. Sage publications.

Patton, M. Q. (2002). Utilization-focused evaluation checklist. Utilization-Focused Evaluation (UFE) Checklist.

Pinto, D. M., & Bin, A. (2024). Use of impact evaluation results in agricultural RD&I organizations: A multi case study. Submitted.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. American journal of evaluation, 29(4), Artigo 4.

Preskill, H., & Boyle, S. (2009). Insights into Evaluation Capacity Building: Motivations, Strategies, Outcomes, and Lessons Learned. Canadian Journal of Program Evaluation, 23(3), 147–174. https://doi.org/10.3138/cjpe.0023.008

Reed, M. S., Ferre, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., & Holden, J. (2021). Evaluating impact from research: A methodological framework. Research Policy. https://doi.org/10.1016/j.respol.2020.104147

Reed, M. S., Gent, S., Seballos, F., Glass, J., Hansda, R., & Fischer-Møller, M. (2022). How can impact strategies be developed that better support universities to address twenty-first-century challenges? Research for All, 6(1). https://www.diva-portal.org/smash/record.jsf?pid=diva2:1713411

Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., & Schubert, A.-L. (2022). Responsible Research Assessment I: Implementing DORA for hiring and promotion in psychology. https://psyarxiv.com/rgh5b/download?format=pdf

Schuijff, M., & Dijkstra, A. M. (2020). Practices of Responsible Research and Innovation: A Review. Science and Engineering Ethics, 26(2), 533–574. https://doi.org/10.1007/s11948-019-00167-3

Stockmann, R., Meyer, W., & Szentmarjay, L. (2022). The Institutionalisation of evaluation in the Americas. Springer Nature.

SPIA. (2020). *SPIA's Approach to Impact Assessment for CGIAR.* Technical Note N.8. Rome: SPIA.

Stockmann, R., Meyer, W., & Taube, L. (2020). The Institutionalisation of evaluation in Europe. Springer Nature.

United Nations Evaluation Group. (2016). Norms and Standards for Evaluation. New York: UNEG.

Van der Most, F. (2010). Use and non-use of research evaluation: A literature review. CIRCLE Paper, 2010/16. http://wp.circle.lu.se/upload/CIRCLE/workingpapers/201016_vanderMost.pdf.

Von Schomberg, R. (2019). Why responsible innovation. International handbook on responsible innovation: a global resource, 12–32.

Weiss, C. H. (1979). The Many Meanings of Research Utilization. Public Administration Review, 39(5). https://doi.org/10.2307/3109916.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? The American Journal of Evaluation, 19(1), 21–33.

Annexes

Annex 1. Survey Questionnaire

Question	Type of answer/options
1. Has management of evaluations been part of your duties?	Yes No
2. What is your age group?	<30 30-40 41-50 51-60 61+
3. What is your gender identity?	Male Female Prefer not to respond Prefer to self-describe [please specify]
4. What is your email address? (Optional) [You will receive a summary of the results if you choose to leave your email address]	Open-Ended Response
5. Where are you based?	East Asia and the Pacific Europe Central Asia Latin America and the Caribbean Middle East and North Africa North America South Asia Sub-Saharan Africa
6. Which organization do you work for?	FAO IFAD WFP UN Agency Multilateral Funds (GEF, GCF…)

Question	Type of answer/options
	International Research Organization
	Donor / Development Bank
	Implementing organization (NGO, Private)
	Government (ministry, institute,…)
	Other [please specify]
7. How many years of experience do you have managing evaluations?	0-3 years
	4-7 years
	8-11 years
	12-29 years
	More than 20 years
8. How much time per month do you allocate to evaluation management?	Less than 10%
	10-29%
	30-49%
	50-74%
	More than 75%
9. What type of evaluations do you usually manage? [select all that applies]	Thematic/Cluster/Sector
	Project/Program
	Corporate
	Region/Country level
	Synthesis
	Impact evaluations/assessments
	Reviews/stocktaking
	Other [please specify]
10. Have you been in charge of developing the evaluation Terms of Reference (ToR)?	Yes
	No
11. Who is usually in charge/responsible of drafting the evaluation ToR?	The entity that commissioned the evaluation
	The evaluation manager
	The consultant or firm
	I don't know
	Other [please specify]

Question	Type of answer/options
12. What is the usual amount of time spent on drafting a ToR for evaluations?	1-2 working days 3-5 working days 6-10 working days 10-30 working days More than one month
13. Who is primarily responsible for the design of the evaluation approach, methodology and the formulation of its main questions?	The entity that commissioned the evaluation The evaluation manager The consultant or firm Other [please specify]
14. Who else participates to/formulates the evaluation questions? [select all that applies]	The entity that commissioned the evaluation The evaluation manager The evaluand team (subject of the evaluation) The consultant or firm Main stakeholders of the evaluand Other [please specify]
15. Is an evaluability assessment usually carried out?	Always Most of the time Sometimes Rarely Never I don't know
16. When is the evaluability assessment usually carried out?	During the evaluand design During the evaluand implementation Before the evaluation During the evaluation inception phase I don't know/NA Other [please specify]
17. Have you ever managed an evaluation with evaluability assessment?	Yes No

Question	Type of answer/options
	Only firms
	Only individual consultants
	Both, but with a preference for firms
	Both, but with a preference for individual consultants
	Both equally
	It depends on the type of evaluation, evaluand and the context [please specify]
18. If you hire individual consultants, who is usually responsible for drafting a ToR for hiring them? [select all that applies]	Myself
	Another evaluation manager/colleague
	Other [please specify]
19. How long is it usually spent on drafting a ToR (or job description) for hiring individual consultants?	1-2 working days
	3-5 working days
	6-10 working days
	More than 10 working days
	I don't know
20. If you hire individual consultants, how long do you usually spend looking for the team of individual consultants?	Less than 2 weeks
	Between 2 weeks and 1 month
	Between 1 and 3 months
	More than 3 months
21. If you hire individual consultants, rate the usual level of difficulty in finding and hiring the right team.	Very difficult
	Somewhat difficult
	Neither easy nor difficult
	Somewhat easy
	Very easy
22. If you hire individual consultants, rate your overall level of satisfaction with hiring individual consultants	Very dissatisfied
	Dissatisfied
	Neither satisfied nor dissatisfied
	Satisfied
	Very Satisfied

Question	Type of answer/options
23. What are the main challenges in finding the right individual consultants/team of consultants? [Rank in order of importance]	Time constraints Limited funding availability Long bureaucratic processes My lack of knowledge of the context and of local consultants Low availability of subject matter experts
24. If you are dissatisfied or very dissatisfied with hiring consultants, could you tell us the main reasons for dissatisfaction?	Open-Ended Response
25. If you hire individual consultants for an evaluation. and the team consists of 2 or more members, who usually serves as team lead?	The evaluator The subject matter expert The evaluation manager It depends on the type of evaluation [please specify]
26. Does your evaluation entity use a rostering process for external consultants? [Rostering process means maintaining a pool of qualified individuals who can be called upon as needed, ensuring that the right people are identified for specific roles or assignments. This process can help facilitate long-term agreements and streamline the management of human resources.]	Yes No I don't know
27. Do you hire only consultants? {if your answer is Yes then we will not ask you about hiring firms in the next section}	Yes No
28. If you hire firms, do you have a long-term agreement with them? [Long-term agreement means here that the firm is hired for a period that lasts more than the length of one evaluation}	Always Most of the time Sometimes Rarely Never
29. If you hire firms, how long do you usually spend looking for the firm?	Less than 2 weeks Between 2 weeks and 1 month Between 1 and 3 months More than 3 months
30. If you hire firms, rate the usual level of difficulty in finding and hiring the right firm	Very difficult Somewhat difficult

Question	Type of answer/options
	Neither easy nor difficult
	Somewhat easy
	Very easy
31. If you hire firms, rate your overall level of satisfaction with the quality of their work	Very dissatisfied
	Dissatisfied
	Neither satisfied nor dissatisfied
	Satisfied
	Very Satisfied
32. What are the main challenges in finding the right firm? [Rank in order of importance]	Time constraints
	Limited funding availability
	Long bureaucratic processes
	My lack of knowledge of the context
	My lack of knowledge of the local firms
	Low availability of firms with specific subject matter expertise
33. What other challenges do you usually encounter?	Open-Ended Response
34. Does your evaluation entity use a rostering process for external consultants? [Rostering process means maintaining a pool of qualified individuals who can be called upon as needed, ensuring that the right people are identified for specific roles or assignments. This process can help facilitate long-term agreements and streamline the management of human resources.]	Yes
	No
	I don't know
35. In your role, do you contribute to the design of the data collection instruments and protocols (such as the interview guide)?	Always
	Most of the time
	Sometimes
	Rarely
	Never
36. During the evaluation, do you participate in interviews, focus groups, and other data collection activities?	Always
	Most of the time
	Sometimes
	Rarely
	Never

Question	Type of answer/options
37. If you participate, do you do so as an observer or do you actively ask questions?	Observer Actively asking It depends [please specify] Not applicable
38. Do you travel to the field during evaluations?	Always Most of the time Sometimes Rarely Never
39. From your perspective, what are the 3 main challenges of data collection?	Open-Ended Response
40. From your perspective, what are the pros and cons with participating in data collection?	Open-Ended Response
41. In terms of data privacy, do you have in place an internal system/process to archive collected data, interview notes and recordings?	Yes No
42. To what extent is Artificial Intelligence (AI) is used in evaluations you are involved in?	AI use is essential in my evaluations as it saves a lot of time It is used only for selective tasks AI is not used at all in our evaluations I don't use it but external consultants are using it
43. If Artificial Intelligence (AI) is used for note-taking and summarizing, is there a quality check performed after the notes are produced?	Always Very frequently Occasionally Rarely Very rarely Never
44. Do you contribute to the original writing of the report? [Here it is about writing not quality assurance of the report]	Always Most of the time Sometimes Rarely Never

Question	Type of answer/options
45. If yes, which parts do you contribute to? [Select all that applies]	Background/context Evaluation methodology Results and key findings Recommendations/conclusions Other [please specify]
46. How does your name appear on the report cover?	Author/co-author Manager Contributor The report is attributed only to the evaluation office itself Other [please specify]
47. Do you submit the draft evaluation report to external peer reviewers for feedback?	Always Most of the time Sometimes Rarely Never
48. Do you agree with the statement: The contribution of external peer reviewers is an added value to the evaluation report"?	Strongly agree Agree Neither agree nor disagree Disagree Strongly disagree
49. Do you submit the draft evaluation report to internal peer reviews for feedback? [By internal peer reviewers, we mean colleagues who are not part of the evaluand team. They can be other evaluation managers, supervisors, or internal subject matter experts.]	Always Most of the time Sometimes Rarely Never
50. Do you agree with the statement: "The contribution of internal peer reviewers is an added value to the evaluation report"?	Strongly agree Agree Neither agree nor disagree Disagree Strongly disagree

Question	Type of answer/options
51. Do you agree with the statement: "As Evaluation Manager, I usually have enough time to properly review the evaluation deliverables (reports, sub-studies, analysis…)?	Strongly agree Agree Neither agree nor disagree Disagree Strongly disagree Not applicable
52. Are your evaluation reports published?	Always Most of the time Sometimes Rarely Never
53. What is the criteria for publishing the evaluation reports?	It depends on the type of evaluation Decided before the evaluation Decided after the evaluation Other [please specify]
54. Please tell us about the usual criteria.	Open-Ended Response
55. How long does it take from validation of the report to its publication?	Less than 1 month Between 2 and 3 months Between 6 and 12 months More than 1 year
56. How long does the development of Management Response usually takes?	Less than 2 weeks Between 2 weeks and 1 month Between 1 and 2 months More than 2 months I don't know
57. Does the management response usually get published?	Always Most of the time Sometimes Rarely Never Not applicable

Question	Type of answer/options
58. If yes, where does the management response get published? [Select all that applies]	In the same document as the evaluation report In a separate document as the evaluation report, but at the same time as the report In a separate document as the evaluation report, at a later date Other [please specify]
59. Who is usually in charge of coordinating management response development?	Evaluation office/manager Evaluand management team Management of the organization Board of the Organization Other [please specify]
60. Is management response developed for all evaluations?	Yes No
61. Who approves/endorses the management response in your organization?	Open-Ended Response
62. Does your organization have a system for tracking status of implementing Management Response?	Yes No I don't know
63. What is the scope of the Management Response tracking system?	Independent evaluations Internal/self-evaluations (not commissioned by funders) MOPAN Reviews Other [please specify]
64. Is the Management Response tracking system publicly available/accessible?	Yes No I don't know
65. If yes, who is in charge of recommendations tracking?	Within the evaluation office/entity/service Top management and/or other divisions of the organization and/or evaluands Other [please specify]

Question	Type of answer/options
66. Who presents the evaluation results to governing bodies and/or donors?	The evaluation manager and/or Supervisors The evaluation team/ the firm Both evaluation manager and the evaluation team None of the above [please specify] Other [please specify]
67. On average, how much time are you given to present the evaluation results to governing bodies?	Less than 1 hour 1-2 hours 3-5 hours More than 5 hours
68. In your organization, how would you rate the use of evaluative evidence for decision making (planning, mid-course correction…):	5 star rating system
69. Please leave your email address if you want to receive a summary of the results.	Open-Ended Response
70. Would you like to share any further thoughts on the topics in this survey?	Open-Ended Response

Annex 2. Overview of EvalforEarth Discussion

Link to the online discussion: https://www.evalforearth.org/discussions/management-matters

Subject of the discussion, copied from the website:

Dear colleagues,

I believe that many would agree with me that the quality of an evaluation isn't solely shaped by the technical expertise of the evaluation team; effective evaluation management is also critical to its success.

Independent evaluation offices within international development agencies employ diverse approaches to managing evaluations: I would like to invite you to reflect and explore how these management approaches influence the credibility and quality of evaluations, including in their connection to use of evidence in decision-making processes.

If you're involved in managing evaluations or interact with independent evaluation functions as an external evaluator/expert, please consider sharing your insights and reflections on the following questions:

- Involvement of evaluation managers: While preserving the independence of evaluations, how should the involvement of an evaluation manager be strategically calibrated across the evaluation phases?
- Role of evaluation managers: To what degree should the role of an evaluation manager encompass active participation as a team member, as opposed to just supervising the evaluation process? What are the advantages and disadvantages of each level of involvement?
- Collaboration with evaluation managers: As an external evaluator, can you share your insights and experiences regarding collaboration with evaluation managers and functions? Has their participation enhanced the relevance and utility of evidence for decision-making processes?

I am looking forward to hearing from you, many thanks in advance!

Please post in the form below or send your contribution to evalforward@evalforward.dgroups.io.

Discussion Summary:

Lal Manavado emphasized the importance of effective evaluation management in facilitating evaluators' work, ensuring relevant data collection, and fostering collaboration. Manavado also highlighted how managers can provide holistic guidance to enhance the evaluation process, especially when evaluators need extensive background information.

Gebril Mahjoub Osman underscored the necessity of preserving the independence of evaluation teams. He provided arguments against the active participation of evaluation managers, due to potential biases and conflicts of interest, and suggested a role in facilitation and support rather than direct involvement.

Vicente Plata stressed the value of effective communication, such as initial and final meetings between evaluation managers and teams, to provide contextual insights and refine conclusions. Plata pointed out that evaluations must balance data with an understanding of the project's broader impact on the actors involved.

Cristian Maneiro recommended that evaluation managers should be supported by evaluation analysts to manage workloads effectively. Maneiro noted that certain evaluation approaches, such as Developmental Evaluation, emphasize a more formative focus. In these cases, the evaluation manager's involvement as an integral part of the evaluated program essential. This approach fosters greater ownership and promotes internal learning within the organization.

Hadera Gebru supported the idea that the roles and responsibilities of evaluation managers should be clearly defined and communicated. Gebru advocated for strategic involvement to ensure high-quality and credible evaluation results, while preserving the independence of evaluators.

Adéléké Oguniyi emphasized the importance of the inception phase as a foundation for successful evaluations. Oguniy highlighted the need for clear communication, setting expectations, and collaborative planning between evaluation managers and external evaluators as key to the process.

Anne Clémence Owen discussed the dual role of evaluation managers in supporting the evaluation process and promoting learning. She pointed that managers' involvement should be clearly defined from the design stage to ensure alignment with evaluation goals and organizational requirements is important.

Musa K. Sanoe noted the significance of proper orientation and clear role definitions for evaluation managers. He emphasized the need for strategic involvement of managers at critical steps to maintain the evaluation’s credibility and independence is very insightful.

Annex 3. Principles and Standards of Independent Evaluation in International Organizations

In international aid and development cooperation, standards have been set to guide how interventions are evaluated. Among several contributions to this topic emerged in recent decades, there are three notable ones that are often cited by the single organization's foundational documents on independent evaluation practices: the OECD-DAC (Development Assistance Committee) principles, the ECG (Evaluation Cooperation Group) good practice standards and the UNEG (United Nations Evaluation Group) norms and standards for evaluation.

The **OECD-DAC** has been publishing several guidelines since the 1990s on development cooperation. In particular, the 2011 **Busan Partnership for Effective Development Co-operation** is a set of common principles for all development actors adopted by representatives of government, multilateral organizations, civil society, private sector, foundations and other actors. It centers on the ownership of development priorities by developing counties, the focus on results, partnerships for development, transparency and shared responsibility. These have been updated in 2019, by defining **six evaluation criteria**–relevance, coherence, effectiveness, efficiency, impact and sustainability–and **two principles** for their use: (1) the criteria should be applied thoughtfully to support high quality, useful evaluation, and (2) the use of the criteria depends on the purpose of the evaluation. The criteria should not be applied mechanistically.

In 2012, the **ECG**, which aims at harmonizing evaluation work among multilateral development banks (MDBs) has published the **Big Book on Evaluation Good Practice Standards.** The document guides the evaluation practice of the five members (African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, InterAmerican Development Bank and World Bank) and contains a detailed discussion on the meaning of **independence** when it comes to evaluation. It states that:

> *"The raison d'être of independence is not for its own sake but to provide for impartial, credible evaluation as a means to help improve the performance of an organization. Four principles should be borne in mind when considering independence*:
>
> 1. *The rationale for independence in its various dimensions is to provide for, and to protect, the impartiality of evaluations and to ensure that the ability of the evaluators to provide credible reports and advice is not compromised.*
> 2. *Independence does not mean isolation, as both operations and evaluation activities are enriched through cross-fertilization of knowledge and experience […]*
> 3. *Independence does not imply any particular approach to evaluation. In particular, independence does not mean that evaluators should focus more on accountability than on learning.*
> 4. *Independence does not mean lack of accountability and responsibility […]. The mechanisms used to ensure adequate levels of accountability for the evaluators may be somewhat different from, and independent of, the mechanisms for the parts of the organization reporting to management."*
> 5.

More recently, and more specifically on the conduct of independent evaluations in international organizations, **UNEG** updated its **Norms and Standards for Evaluation** publication. UNEG is an inter-agency professional network created in 1984 that brings together the evaluation units of the UN system. Its mission

is to support the strengthening and harmonization of evaluation practices among its more than 60 full members and observers. The 2016 Norms and Standards for Evaluation updates the 2005 version, a foundational document intended for application for all United Nations evaluations. The update sheds lights onto evaluations in the UN system in the era of the 2030 Sustainable Development Agenda.

The 2016 document defines **evaluation** as follows:

> *"An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability. An evaluation should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making processes of organizations and stakeholders."*

The document defines **norms** for evaluation that should guide the establishment and practice of the organizations' independent evaluation function. The first ten cover general norms for evaluation: (1) **Internationally agreed principles, goals and targets** are to be upheld and promoted by evaluation managers and evaluators; (2) **Utility** of evaluation results must be ensured by informing decisions and actions; (3) **Credibility** must be grounded on independence, impartiality and a rigorous methodology; (4) **Independence**, must be both behavioral and organizational; (5) **Impartiality** should be at all stages of the evaluation process; (6) **Ethics**, so that evaluation upholds the highest standards of integrity and respect for the beliefs, manners and customs of the social and cultural environment; for human rights and gender equality; and for the 'do no harm' principle for humanitarian assistance; (7) **Transparency** must be pursued to enhance stakeholder ownership and increase public accountability; (8) **Human rights and gender equality** need to be integrated into all stages of an evaluation; (9) **National evaluation capacities** should be supported upon the request of member states; (10) **Professionalism** should contribute towards the credibility of evaluators, evaluation managers and evaluation heads, as well as the evaluation function. Four further norms concern the Institutional norms for evaluation in the UN system; (11) The **enabling environment** must include an organizational culture that values evaluation as a basis for accountability, learning and evidence-based decision-making; (12) An explicit **evaluation policy** must be in established; (13) the **responsibility for the evaluation function** establishment lies with the governing and/or executive bodies of the organization; and (14) **Evaluation use and follow-up** must be promoted using an interactive process that involves all stakeholders.

To aid the implementation of the 14 norms, the document further defines 24 **standards** for evaluation, organized in 5 groups:

1. **Institutional Framework**
   STANDARD 1.1 Institutional framework for evaluation: The organization should have an adequate institutional framework for the effective management of its evaluation function.

   STANDARD 1.2 Evaluation policy: Organizations should establish an evaluation policy that is periodically reviewed and updated to support the evaluation function's increased adherence to the UNEG Norms and Standards for Evaluation.

   STANDARD 1.3 Evaluation plan and reporting: Evaluations should have a mechanism to inform the governing body and/or management on the evaluation plan and on the progress made in plan implementation.

STANDARD 1.4 Management response and follow up: The organization should ensure that appropriate mechanisms are in place to ensure that management responds to evaluation recommendations. The mechanisms should outline concrete actions to be undertaken in the management response and in the follow-up to recommendation implementation.

STANDARD 1.5 Disclosure policy: The organization should have an explicit disclosure policy for evaluations. To bolster the organization's public accountability, key evaluation products (including annual reports, evaluation plans, terms of reference, evaluation reports and management responses) should be publicly accessible.

2. **Management of the Evaluation Function**

STANDARD 2.1 Head of evaluation: The head of evaluation has the primary responsibility for ensuring that UNEG Norms and Standards for Evaluation are upheld, that the evaluation function is fully operational and duly independent, and that evaluation work is conducted according to the highest professional standards.

STANDARD 2.2 Evaluation guidelines: The head of evaluation is responsible for ensuring the provision of appropriate evaluation guidelines.

STANDARD 2.3 Responsiveness of the evaluation function: The head of evaluation should provide global leadership, standard setting and oversight of the evaluation function to ensure that it dynamically adapts to new developments and changing internal and external needs.

3. **Evaluation Competencies**

STANDARD 3.1 Competencies: Individuals engaged in designing, conducting and managing evaluation activities should possess the core competencies required for their role in the evaluation process.

STANDARD 3.2 Ethics: All those engaged in designing, conducting and managing evaluations should conform to agreed ethical standards to ensure overall credibility and the responsible use of power and resources.

4. **Conduct of Evaluations**

STANDARD 4.1 Timeliness and intentionality: Evaluations should be designed to ensure that they provide timely, valid and reliable information that will be relevant to the subject being assessed and should clearly identify the underlying intentionality.

STANDARD 4.2 Evaluability assessment: An assessment of evaluability should be undertaken as an initial step to increase the likelihood that an evaluation will provide timely and credible information for decision-making.

STANDARD 4.3 Terms of reference: The terms of reference should provide the evaluation purpose, scope, design and plan.

STANDARD 4.4 Evaluation scope and objectives: Evaluation scope and objectives should follow from the evaluation purpose and should be realistic and achievable considering resources available and the information that can be collected.

STANDARD 4.5 Methodology: Evaluation methodologies must be sufficiently rigorous such that the evaluation responds to the scope and objectives, is designed to answer evaluation questions and leads to a complete, fair and unbiased assessment.

STANDARD 4.6 Stakeholder engagement and reference groups: Inclusive and diverse stakeholder engagement in the planning, design, conduct and follow-up of evaluations is critical to ensure ownership, relevance, credibility and the use of evaluation. Reference groups and other stakeholder engagement mechanisms should be designed for this purpose.

STANDARD 4.7 Human rights-based approach and gender mainstreaming strategy: The evaluation design should include considerations of the extent to which the United Nations system's commitment to the human-rights based approach and gender mainstreaming strategy was incorporated in the design of the evaluation subject.

STANDARD 4.8 Selection and composition of evaluation teams: The evaluation team should be selected through an open and transparent process, considering the required competencies, diversity in perspectives and accessibility to the local population. The core members of the team should be experienced evaluators.

STANDARD 4.9 Evaluation report and products: The final evaluation report should be logically structured and contain evidence-based findings, conclusions, and recommendations. The products emanating from evaluations should be designed to the needs of its intended users.

STANDARD 4.10 Recommendations: Recommendations should be firmly based on evidence and analysis, clear, results-oriented and realistic in terms of implementation.

STANDARD 4.11 Communication and dissemination: Communication and dissemination are integral and essential parts of evaluations. Evaluation functions should have an effective strategy for communication and dissemination that is focused on enhancing evaluation use.

5. **Quality**

STANDARD 5.1 Quality assurance system: The head of evaluation should ensure that there is an appropriate quality assurance system.

STANDARD 5.2 Quality control of the evaluation design: Quality should be controlled during the design stage of evaluation.

STANDARD 5.3 Quality control at the final stage of evaluation: Quality should be controlled during the final stage of evaluation.

Annex 4. Evaluation Mapping Across Peer Organizations

1. Background and Objective

The purpose of the mapping was to gather insights into the output of evaluation functions of international organizations with missions similar to CGIAR and responsive to MOPAN. The result aimed to inform the development of the Multi-Year Evaluation Plan (MYEP) 2025-27 to best align to industry standards and steer expectations from clients. In this desk review, target organizations were first identified, and then a sample of their recent evaluation reports against a few characteristics were analyzed, including which types of evaluations are conducted, the time required to publish reports, the number of countries visited, and the size of the teams involved in each evaluation.

2. Data and Coverage

A sample of evaluations from nine external organizations were selected, considering at least one evaluation report from recent years for each of the themes, stated by the organization themselves. One or more report for each theme was selected randomly among the publicly available ones. Two evaluation reports from the IAES Evaluation Function of CGIAR were selected to reach a final number of 100 evaluations for analysis (see Table A1.) Reports were downloaded from the websites of organizations' evaluation functions in November 2023 and August 2024. Regarding the choice of sampled organizations, the Rome-Based UN Agencies, Development Banks, and other organizations whose missions and themes align with those of CGIAR were chosen.

Table A1. Distribution of evaluation reports across the sample

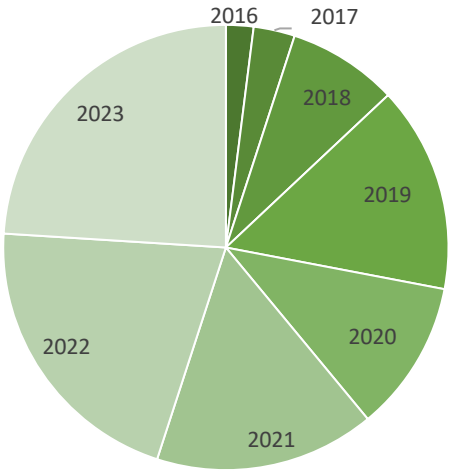
Organization	Number of included evaluation reports	Organization-specific themes covered in the sample	
African Development Bank (AfDB)	10	• Corporate evaluation (1) • Country strategy and program evaluation (1) • Evaluation Synthesis (1) • Impact evaluation (1) • Other knowledge product (1) •	• Project cluster evaluation (1) • Project completion report and expanded supervision report Validation synthesis (1) • Regional integration strategy evaluation (1) • Sector evaluation (1) • Thematic evaluation (1)
Food and Agriculture Organization of the United Nations (FAO)	12	• Country (2) • Joint evaluation (1) • Programme (2)	• Project (3) • Synthesis (2) • Thematic evaluation (2)
Global Environment Facility (GEF)	9	• Country (2) • Institutional (3) • Performance (2) • Thematic (2)	

Organization	Number of included evaluation reports	Organization-specific themes covered in the sample	
Green Climate Fund (GCF)	11	• Performance review (2) • Portfolio (2) • Programmatic (5) • Thematic (2)	
International Fund for Agricultural Development (IFAD)	12	• Annual report (1) • Corporate level evaluation (1) • Corporate-level and thematic evaluations (2) • Country strategy and programmes (2) • Evaluation Synthesis Reports (2) •	• Impact evaluations (1) • Project cluster evaluation (1) • Project evaluations (1) • Regional evaluation (1)
UN Women	12	• Cluster evaluation (2) • Corporate (organizational performance evaluation) (1) • Country-level evaluation (outcome level) (2) •	• Programme evaluation (3) • Regional evaluation (1) • Strategy/policy evaluation (1) • Thematic evaluation (2)
United Nations Development Programme (UNDP)	10	• Decentralized country programme Evaluation (1) • Independent country programme Evaluation (3) • Outcome (2) • Portfolio evaluation (1) • Project (2) • Thematic (1)	
World Food Programme (WFP)	12	• Country portfolio and country strategic plan evaluations (2) • Humanitarian emergency response evaluations (1) • Impact evaluations (2) • Operations and activity evaluations (3) • Policy evaluations (1) • Strategic evaluations (1) • Thematic and transfer modality evaluations (2)	
The World Bank	10	• Country focused validations (1) • Country focused evaluations (1) • IEG annual reports (1) • Major evaluations (1) • MAR validation Report (1) • Meso evaluations (1) • Portfolio (1) • Project level evaluations (2)	

Organization	Number of included evaluation reports	Organization-specific themes covered in the sample	
		• Synthesis papers (1)	
CGIAR IAES	2	• Corporate (2)	

As shown in Figure A1, the included evaluation reports were published between 2016 and 2023, with 72% of them published in the last five years (since 2019). For this selection, mainly reports from recent years were included, however in some cases, older reports were assessed, due to lack of availability. A total of 16 out of the 100 were included and mentioned in a MOPAN review of the organization's evaluation function.

Figure A1. Distribution of evaluations included in the review by year of publication



3. Methodology and Results

Once a satisfactory sample of evaluation reports that spanned across all the themes stated by the organizations were selected, these reports were systematically analyzed across several characteristics, whenever such characteristics were stated in the evaluation reports. The types of evaluations conducted were first looked at, re-coding the stated themes into a harmonized categorization that could be applied across the different entities. Then the duration of evaluations, the size of the teams employed, and the number of countries visited were analyzed.

The last column of Table A1. Distribution of evaluation reports across the sampleclearly shows that the way organizations categorize the types of evaluations varies greatly across the entities. To gain a better picture of evaluation types, the different evaluation categories were recorded into a list of 21 harmonized types, as shown in Table A2.

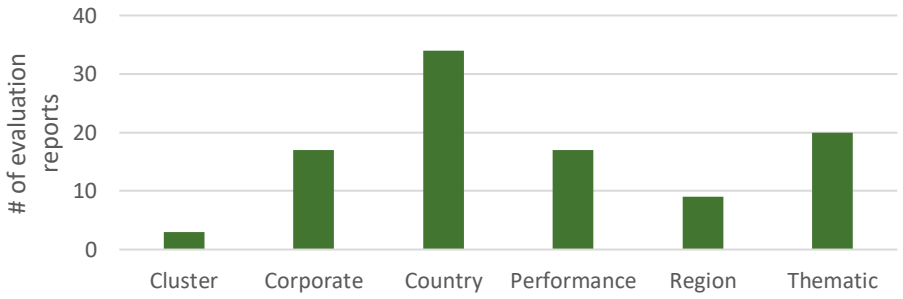
Table A2 . Harmonized categorization of evaluation types across organizations

Harmonized evaluation types	Shortened list
1. **Cluster evaluation**	1. Cluster

Harmonized evaluation types	Shortened list
2. **Corporate - function/system/process**	2. Corporate
3. **Country - impact**	3. Country
4. **Country - portfolio**	4. Performance
5. **Country - programme**	5. Region
6. **Country - project**	6. Thematic
7. **Country - strategy**	
8. **Evaluative reviews**	
9. **Mid-term**	
10. **Performance**	
11. **Performance - review**	
12. **Process**	
13. **Region**	
14. **Region - programme/project**	
15. **Reviews of evaluation**	
16. **Thematic**	
17. **Thematic - approach**	
18. **Thematic - contribution**	
19. **Thematic - programme**	
20. **Thematic - SDG**	
21. **Thematic - strategy**	

When organized according to the harmonized list, our sample is distributed as shown in Figure A2.

Figure A2. Distribution of evaluations by harmonized type of publication (shortened list)



For what concerns the modalities in which evaluations were conducted, whenever available, the approximate duration of the evaluations was collected. The length in 78 cases was determined, and among these, the average duration has been just under a year (11 months), ranging from a minimum of one month to a maximum of two years and two months. Figure A3 shows the distribution. Figure A4 shows the average length of evaluation by harmonized evaluation types.

Figure A3. Distribution of evaluations by average approximate duration

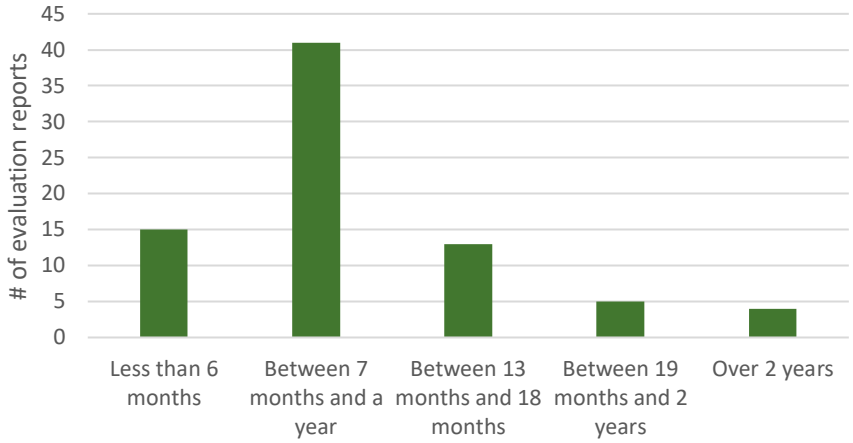
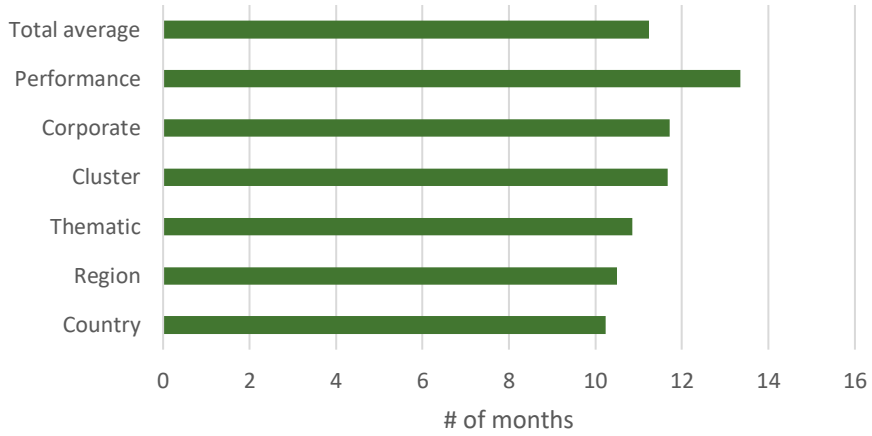


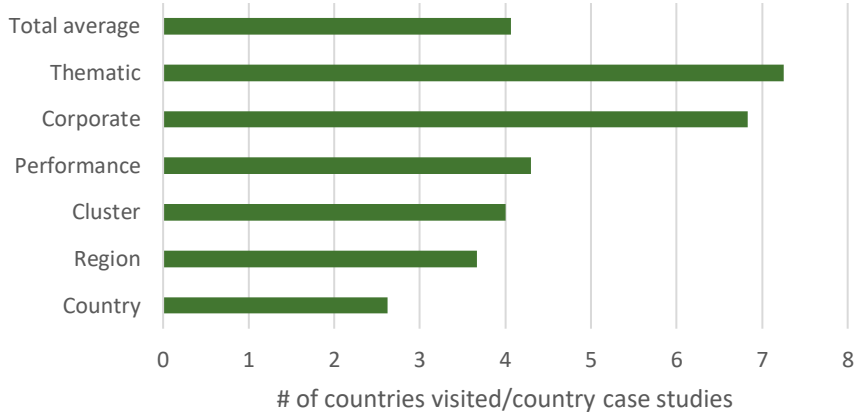
Figure A4. Distribution of evaluations by average approximate duration (shortened list)



Whenever available, the number of countries visited during the evaluation period were compiled. Since most of the evaluation reports have been conducted during the COVID-19 years, in many cases no countries could be visited, therefore this definition was expanded to number of country case studies included, to represent the number of countries that would have been visited, if it had not been for the travel disruption of the global pandemic. This information for 56 of the reports were able to be retrieved. While the average was 3.7 countries per evaluation study, most included a visit or case study for one country, and the reports ranged all the way up to a maximum of 43 countries. Figure A5 shows the distribution, while Figure A6 shows the average number by harmonized evaluation types.

Figure A5. Distribution of evaluations by number of countries visited/country cases included

Figure A6. Average number of countries visited/country case studies included by harmonized evaluation type (shortened list)



The final characteristic of evaluation modality compiled, whenever available, was the approximate number of team members employed per evaluation. This information was available for 93 of the 100 sampled evaluations and ranged from a minimum of one to a maximum of 42, with an average of eight team members. Figure A7 shows the distribution across the sample, and Figure A8 shows the breakdown by harmonized evaluation type.

Figure A7. Distribution of evaluations by number of team members

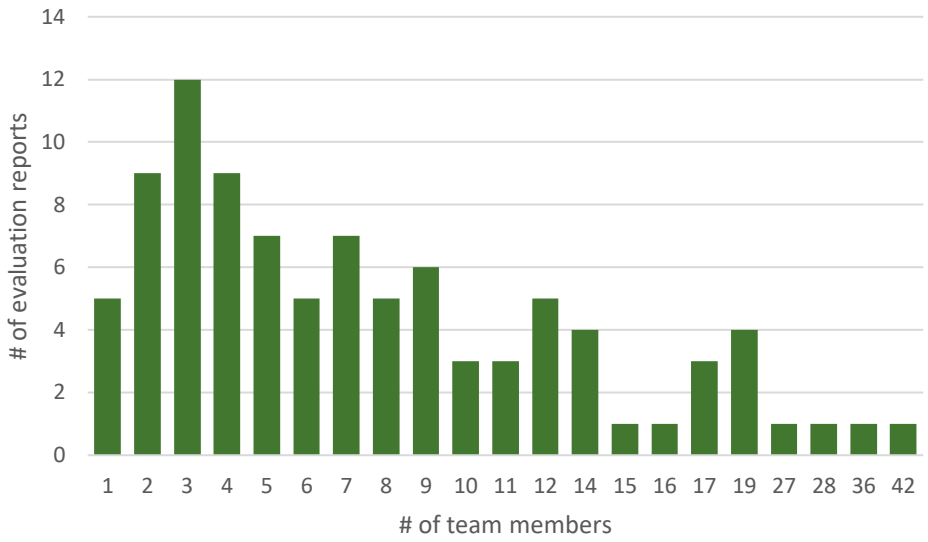
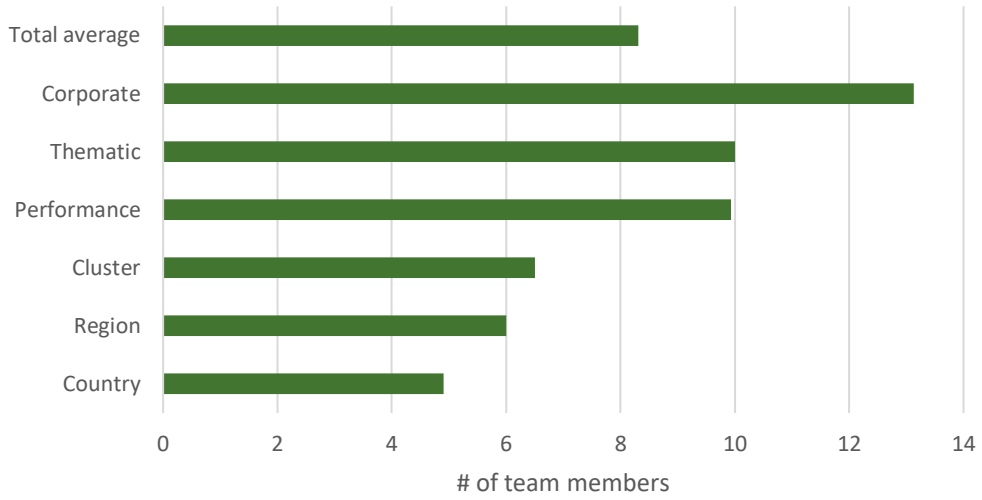


Figure A8. Average number of team members by harmonized evaluation type (shortened list)



Independent Advisory and Evaluation Service
Alliance of Bioversity International and CIAT
Via di San Domenico, 1 00153 Rome, Italy
IAES@cgiar.org