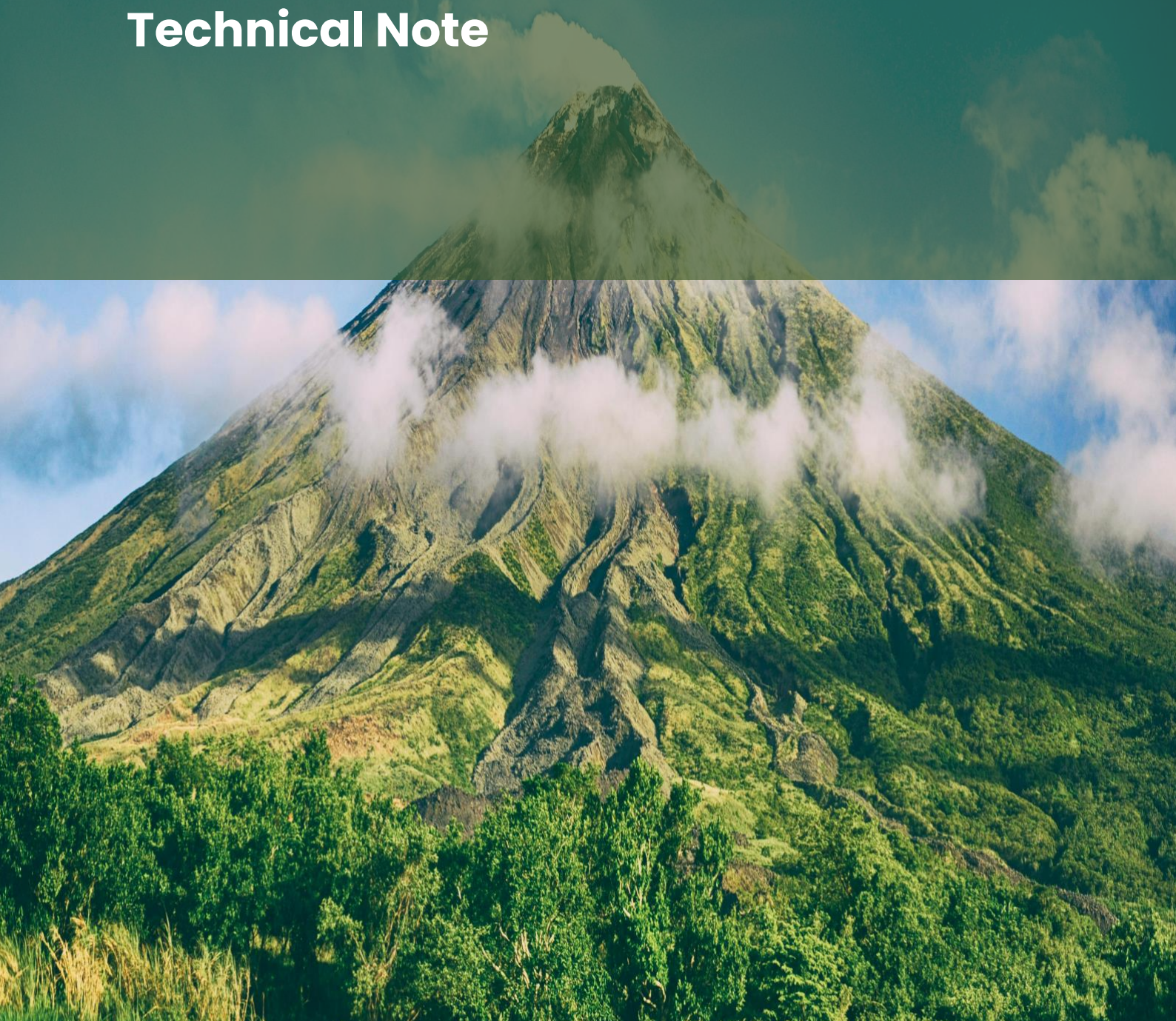




Independent
Advisory and
Evaluation
Service

Evaluation of Scaling Efforts in CGIAR

Technical Note





Independent
Advisory and
Evaluation
Service

Correct citation: CGIAR Independent Advisory and Evaluation Service (IAES). 2026. Evaluation of Scaling Efforts in CGIAR. Technical Note. Rome: IAES, Evaluation Function. <https://iaes.cgiar.org/evaluation>

Cover image: Archie Binamira. Tabaco City, Bicol, Philippines.

Evaluation of Scaling Efforts in CGIAR

Technical Note

March 2026

Acknowledgments

This Note was prepared by independent consultants John Gargani and Hezekiah Agwara, together with IAES Evaluation Analyst Matt Armstrong, under the guidance of, and with substantive contribution from, Senior Evaluation Manager Ibtissem Jouini. Overall direction and valuable feedback were provided by IAES Director Allison Grove Smith and Evaluation Function Lead Svetlana Negroustoueva.

The Note benefited from external peer reviews by Amy Jersild and Amitaksha Nag. Reviews by primary users were provided by external evaluators Keri Culver and Natascia Palmieri, as well as evaluation analysts Gaia Gullotta and George Theuri. Additional review was provided by Evaluation Reference Group member Ravinder Kumar.

Substantive feedback and inputs were also contributed by Marc Schut, Senior Innovation Portfolio Management Advisor, CGIAR Office of the Chief Scientist, Portfolio Performance Team. Timothy Krupnik, Director of the CGIAR Scaling for Impact (S4I) Program, further enriched the Note with valuable insights and updates. The Note also benefited from expert peer review by two members of the scaling community of practice, Johannes F. Linn and Richard Kohl.

Hanna Ewell, Research Specialist Climate Action at the Alliance of Bioversity and CIAT, provided detailed and constructive feedback.

IAES staff provided essential support throughout the process, ensuring the smooth coordination of technical and administrative activities. Copy-editing was provided by Kelly Blank, and layout and design by Federica Bottamedi.

Soliciting Input and Feedback

The IAES would like to receive feedback about learning from the roll-out and application of this first version of the Technical Note and invites users to contact the Independent Advisory and Evaluation Service (IAES)/Evaluation Function (IAES-Evaluation@cgiar.org) as custodian of the document. The next version will continue building on direct experiences and evolving industry standards on how to evaluate scaling efforts.

Title	Evaluation of Scaling Efforts in CGIAR: A Technical Note
Purpose	Help evaluators undertake process and performance evaluations of scaling efforts in CGIAR.
Audience	Evaluators, evaluation managers and commissioners involved in evaluating scaling in CGIAR as well as CGIAR colleagues conducting assessments of their own scaling efforts. Those responsible for providing input into evaluations in the CGIAR system; users in other agricultural research-for-development (AR4D) contexts may find this document useful too.
Framework and Policy Reference	This guidance supports the CGIAR Evaluation Framework and the CGIAR Evaluation Policy (2022) and should be read in conjunction with other evaluation-related guidelines.
Contact	For queries and feedback about learning from the roll-out and application contact the Evaluation Function within the IAES of CGIAR at IAES-Evaluations@cgiar.org .

Contents

Table of Acronyms	vii
Overview	1
1 Introduction	2
1.1 CGIAR's Diverse Scaling Efforts.....	3
1.2 Scope of the Technical Note.....	4
2 Clarifying Scaling Terms and Concepts	6
2.1 What is an Innovation?	6
2.2 What is Scaling?.....	7
2.3 What is a Scaling Strategy and What are Scaling Pathways?	7
3 Process and Performance Evaluation of Scaling: A New Area of Practice	8
3.1 Familiar Ground: Evaluating Programs.....	8
3.2 New Territory: The Logic of Process and Performance Evaluations of Scaling	9
3.3 Learning From Experience and Attending to Power	10
4 Scaling in the CGIAR Context and Implications for Evaluations	11
5 Framework for Evaluating Scaling Efforts	18
5.1 How to Prepare	18
5.2 Criteria, Principles and Guiding Questions.....	19
References	23
Annex 1. CGIAR MELIA Glossary v4.1 (September 2025)	26
Annex 2. Additional Information about Terms and Concepts	29
Annex 3. Innovation Readiness and Use Scores	32
Annex 4. Scaling Theory of Change and Stage Transition Model	37

List of Tables

Table 1. Scaling evaluation criteria and principles	2
Table 2. How applying the logic of process and performance evaluation differs for scaling vs. non scaling programs.....	9
Table 3. Key steps to preparing for the evaluation	19
Table 4. Scaling evaluation criteria, principles and questions.....	20
Table 5. CGIAR MELIA Glossary v4.1 (September 2025) for scaling.....	26
Table 6. Innovation Readiness Scores and their interpretation.....	32
Table 7. Innovation Use Scores and their interpretation	33
Table 8. Scaling metrics for innovations, enablers, bundles, and packages.....	35
Table 9. Breakdown of a CGIAR innovation package and metrics for the Joint Village Participatory Land Use Planning (JVLUP)	36

List of Figures

Figure 1. Scaling dynamics: how impact changes as adoption grows.....	15
Figure 2. Scaling process.....	16
Figure 3. Mapping the scaling process	16
Figure 4. Snapshot of the CGIAR W1/2 innovation pipeline (2024).....	18
Figure 5. ToC of an innovation.....	37
Figure 6. Scaling ToC, scaling an innovation and scaling impact	37
Figure 7. Considering context in a scaling ToC.....	38
Figure 8. The underlying stage transition matrix.....	39

List of Boxes

Box 1. Examples of CGIAR's scaling efforts at 2025–30 Portfolio level	4
Box 2. Key definitions of process and performance evaluations vs. impact assessment in CGIAR.....	5
Box 3. The Green Revolution.....	10
Box 4. Suggested structure for scaling ambition statements	12
Box 5. One way a scaling thesis may be structured.....	13

Table of Acronyms

AR4D	Agricultural Research-for-Development
B4T	Breeding for Tomorrow
CGIAR	Consultative Group on International Agricultural Research
CoP	Community of Practice
CIR	Component Importance Rank
CIW	Component Importance Weight
DAC	Development Assistance Committee
EA	Evaluability Assessment
IAES	Independent Advisory Evaluation Service
IFA	CGIAR Integration Framework Agreement 2022
IPSR	Innovation Packaging and Scaling Readiness
IR	Innovation Readiness
IISDC	Independent Science Development Council
IU	Innovation Use
JVLUP	Joint Village Participatory Land Used Planning
MELIA	Monitoring, Evaluation, Learning and Impact Assessment
MELIA-F	Monitoring, Evaluation, Learning, Impact Assessment, and Foresight
MOPAN	Multilateral Organization Performance Assessment Network
NPS	National Policies and Strategies
QoR4D	Quality of Research for Development
R4D	Research-for-Development
RIIs	Regional Integrated Initiatives
RTB	Roots, Tubers, and Bananas
S4I	Scaling for Impact Program
SPIA	Standing Panel on Impact Assessment
SPP	Scaling Potential of a Package
SRC	Scaling Readiness Credibility
SRI	Scaling Readiness of an Innovation
SRP	Scaling Readiness of a Package
ToC	Theory of Change
ToR	Terms of Reference

Overview

This Note provides guidance on conducting performance and process evaluations of CGIAR's scaling efforts through its 2025–30 Portfolio. Across the Portfolio, Science Programs pursue diverse scaling projects aligned with their research strengths and CGIAR's Impact Areas. Accelerators support these efforts through specialized expertise in gender equity and inclusion, digital transformation, and capacity sharing. And the Scaling for Impact Program (S4I) enables Science Programs to bundle complementary innovations, integrate supporting components, and prepare them for scaling through external partnerships.

Through this Portfolio structure, CGIAR positions itself as a catalyst for scaling, supporting stakeholders, partners, and funders in collaboratively scaling effective innovations. Accordingly, the primary purpose of the evaluations considered here is to support CGIAR in advancing innovations along its innovation development pipeline, from discovery science to external collaborations, to achieve the 'best' impact from scaling-ready innovations. This Note does not address CGIAR's direct engagement in other scaling efforts, such as bilateral programs, though its guidance may inform their evaluation.

Section 1 introduces the evaluation of scaling efforts and defines the scope of the Note. It presents four illustrative examples of CGIAR's diverse scaling approaches to orient evaluators to the context in which they will work. In whole or in part, these evaluations will assess scaling efforts themselves at the level of Programs, Accelerators, or portfolios and their components. This contrasts with more traditional program evaluations, which assess the interventions or innovations being scaled rather than the scaling effort itself. The Note is limited to process and performance evaluation and does not address impact evaluation.

Section 2 presents key terms and concepts. Definitions are drawn mainly from the Monitoring, Evaluation, Learning and Impact Assessment (MELIA) [Glossary v4.1](#) (2025) and the [CGIAR 2030 Research and Innovation Strategy](#). Where needed, the authors explain why broader or adapted definitions are used to reflect the diversity of CGIAR's scaling efforts and to connect with the wider evaluation literature on scaling.

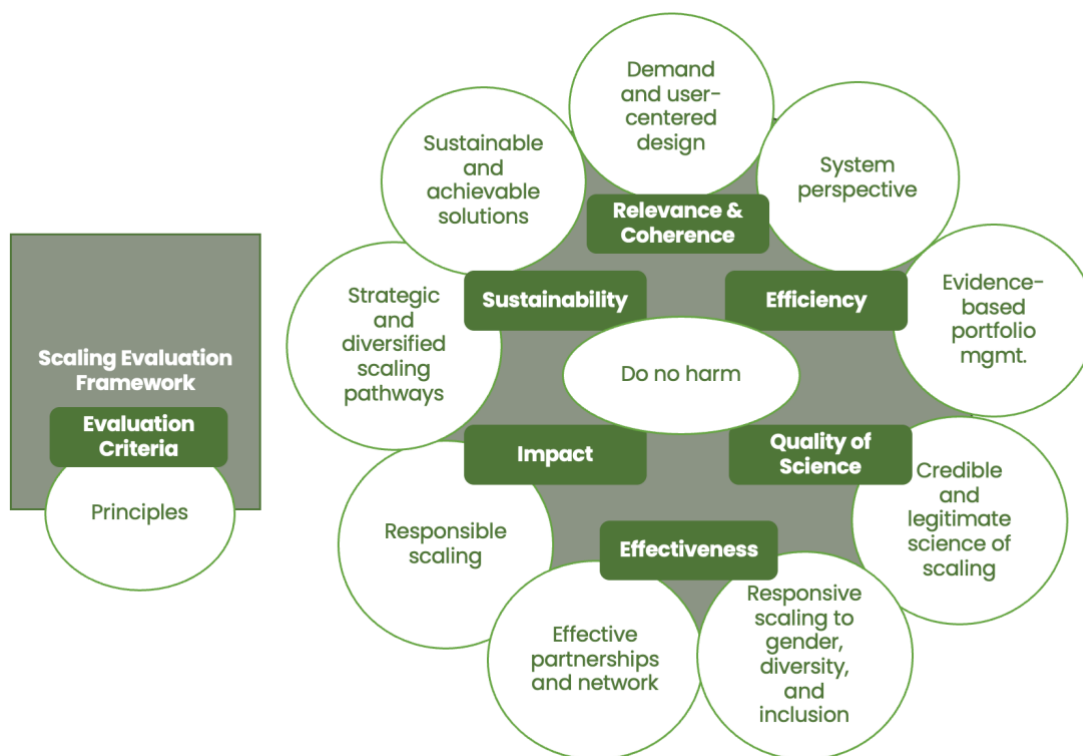
Section 3 focuses on the process and performance evaluation of scaling. The section guides readers from familiar approaches used in Program evaluation (such as expected results, evaluation criteria, evidence, and standards) toward the specific logic of evaluating scaling efforts. Practical examples show how process and performance evaluation differ for scaling compared to non-scaling programs. Then, building on past CGIAR experience, it shows why understanding decision-making and power dynamics is necessary in scaling processes.

Section 4 introduces key scaling concepts, including scaling ambition; the distinction between innovations, bundles, and packages; choices about what to scale using evaluative rubrics; scaling thesis and theory of change (ToC); scaling effects; and mapping the scaling process.

Section 5 concludes with a framework for evaluating scaling efforts in CGIAR, supported by a set of guiding questions. The framework organizes guiding questions around evaluation criteria and principles, it aligns with the CGIAR Evaluation Framework and Policy (2022), using OECD Development Assistance Committee (DAC) criteria adapted to CGIAR's scaling context, and was developed jointly by CGIAR scaling teams and the IAES Evaluation Function.

The annexes are an integral part of the Note. They contain additional detail to keep the main text clear and readable while allowing readers to choose how deeply they wish to engage with the Note.

Table 1. Evaluation criteria and principles for evaluating scaling efforts



1 Introduction

The purpose of this Technical Note is to guide evaluators conduct process and performance evaluations of CGIAR’s scaling efforts (CGIAR Programs, activities, Initiatives, projects), an important task key to CGIAR’s [CGIAR 2030 Research and Innovation Strategy](#). CGIAR has a long-standing commitment to scaling the public benefits of its innovations and has recently strengthened that commitment, increasing the need for systematic evaluation of scaling efforts. This Note outlines how to conduct these evaluations while recognizing that scaling is a rapidly evolving field of science and practice. In development and research for development, methods for evaluating scaling continue to evolve, and the guidance presented here will be tested and refined through practice. IAES will update the Note to reflect what CGIAR learns as it gains experience evaluating scaling.¹

While scaling is central to development effectiveness, existing international standards and tools, such as those of OECD- Development Assistance Committee (DAC) and Multilateral Organization Performance Assessment Network (MOPAN), do not adequately address scaling. They provide limited guidance on how to assess an organization’s approach to scaling or its performance in scaling projects and programs, as highlighted in recent work by the 2025 Scaling Community of Practice (CoP). This Note aims to support stakeholders involved in evaluating scaling efforts in CGIAR by bringing together key concepts, linking them

¹ The first update of the Note will follow the 2027 midline evaluations of the 2025–30 Portfolio.

to CGIAR activities and terminology, and offering a framework for evaluating scaling efforts, supported by guiding questions.

The journey towards the development of this Note began with a desk review and consultations with key CGIAR internal stakeholders, in particular, by consultations and interviews undertaken by IAES Evaluation Function during the Scaling Week 2024 in Nairobi. This was followed by a workshop in Rome on 6 February 2025, where key concepts and steps were presented to a group of CGIAR colleagues to gather insights and diverse perspectives on scaling in CGIAR. The Note was then refined through several rounds of feedback and peer reviews (see the Acknowledgements section). An online [discussion](#) on the evaluation of scaling, initiated by the co-author of this Note on the [EvalforEarth](#) CoP, further enriched the content.

1.1 CGIAR's Diverse Scaling Efforts

Scaling efforts in CGIAR include center level projects and Programs and System level portfolios implemented by the CGIAR System Office.² At system level, CGIAR is currently implementing its [2025-30 Portfolio](#), comprising eight Science Programs, four Accelerators, and the Scaling for Impact (S4I) Program.³ The 2025 Portfolio inception reports describe diverse approaches to scaling a wide range of innovations to impact in different contexts linked to CGIAR five Impact Areas : (1) Nutrition, Health and Food Security; (2) Poverty Reduction, Livelihoods and Jobs; (3) Environmental Health and Biodiversity; (4) Gender Equality Youth and Social Inclusion; and (5) Climate Adaptation and Mitigation. To convey the diversity of scaling efforts that may be evaluated, this section provides snapshots (see Box 1) of the S4I program,⁴ two Science Programs, and one Accelerator (see also other examples in the section about CGIAR's approaches to scaling).

² CGIAR has three funding windows: Window (W) 1: Pooled funding for system-wide CGIAR priorities and core research. W2: Previously Program-targeted pooled funding (now mostly merged into W1) and W3: Donor-directed funding for specific projects at specific centers.

³ Details of the Portfolio are available at [CGIAR Research Portfolio 2025-30](#) and the narrative at [CGIAR Portfolio Narrative 2025-30](#). Thirteen Programs and Accelerators, namely: Breeding for Tomorrow, Genebanks, Policy Innovations, Better Diets and Nutrition, Gender Equality and Inclusion, Multifunctional Landscapes, Food Frontiers and Security, Scaling for Impact, Digital Transformation, Sustainable Farming, Climate Action, Capacity Sharing and, Sustainable Animal and Aquatic Foods.

⁴ S4I builds on CGIAR's 2022-24 [Regional Integrated Initiatives \(RIIs\)](#), which operated in Africa (East and Southern Africa; Central and West Asia and North Africa), Asia (South Asia; Southeast Asia), and Latin America and the Caribbean along with the [National Policies and Strategies](#) (NPS) Initiative. While these Initiatives combined research, innovation and scaling, S4I focuses primarily on the latter.

Box 1. Examples of CGIAR’s scaling efforts at 2025–30 Portfolio level



Snapshot 1: Integrating CGIAR’s Scaling Efforts

S4I is CGIAR’s first Program fully dedicated to scaling Food, Land and Water (FLW) system innovations. It is CGIAR’s focal point for scaling efforts and research to help bridge the gap between innovation development and adoption/use by farmers, governments, and other FLW system actors. The Program’s premise is that a systematic, evidence-based approaches to scaling will help more innovations reach an appropriate scale, thereby producing larger and qualitatively better impacts. The program comprises five interconnected areas of work (AoWs) that establish a framework for driving substantial and inclusive impact.



Snapshot 2: Scaling New Crop Varieties Through Markets

The Breeding for Tomorrow (B4T) Program aims to develop and deliver new crop varieties that promote productive, resilient, inclusive, and sustainable food systems. The Program scaling efforts use a demand-driven approach to achieve development objectives. The premise is straightforward: the same number of genetically innovative crop varieties can create greater impact if more farmers adopt them, and farmers are more likely to adopt new varieties if consumers want them. Consequently, the program will gather evidence of what markets demand and then develop nutritious, climate-resistant crop varieties that meet the demand.



Snapshot 3: Scaling Nutrition Through Systems Transformation

The Better Diets and Nutrition Program works to transform food systems in low- and middle-income countries so that three billion people can benefit from sustainable, healthy diets. The program works from the bottom-up to co-design farm-level innovations and strategies to change consumer behavior. Concurrently, it works from the top down, and closely with policy makers, to design policies and technical interventions. This makes it possible to intervene on multiple leverage points in a food system, by engaging diverse public- and private-sector actors, such as smallholders, consumers, policy makers, and scientists.



Snapshot 4: Scaling Technology to Create Network Effects

The Capacity Sharing Accelerator (CapSha) strengthens the capacity of participating partners and CGIAR to transform FLW systems while advancing gender equality, women’s empowerment, and youth inclusion. It proposed an innovative digital tool, CapSha Marketplace, to connect partners that want to develop capacity in a particular domain with those who can help them develop it. The Accelerator’s proposal focuses on a network of partners composed of CGIAR Programs and members of the National Agricultural Research and Innovation Systems. CapSha Marketplace has the potential to accelerate the impact of capacity sharing through network effects. The more partners that develop a given capacity, the more opportunities there are for other partners to develop it, and this has the potential to increase the rate at which partners learn from each other. As each partner’s capacity to transform systems grows, the more likely their collective efforts will transform them.

CGIAR’s scaling efforts extend beyond the 2025–30 Portfolio and will continue to evolve in response to emerging needs and circumstances⁵. As a result, evaluators may encounter novel scaling efforts. This Note primarily serves the scaling Initiatives described in the 2025–30 Portfolio and may require adaptation to support evaluation across the full range of current and future scaling contexts.

1.2 Scope of the Technical Note

Innovations are routinely assessed to inform decisions about what to scale, and they may also be evaluated at different scales. This Note, however, examines how scaling itself can be evaluated, rather than the intervention or innovation being scaled. Scaling itself can therefore be treated as an intervention, with

⁵ Information collected through consultations with CGIAR stakeholders.

distinct mechanisms and consequences.⁶ Scaling actors should evaluate before, during, and after scaling. Throughout this document, the term scaling actors refers collectively to innovators, investors, and other key players involved in enabling and supporting scaling.

The scope of this Note is to provide guidance, language and frameworks on the emerging field of evaluating scaling to teams undertaking process and performance evaluations (see definition in Box 2) throughout the scaling process, aligned with the scope of [CGIAR's Evaluation Framework and Policy](#) (2022) and the [mandate of IAES Evaluation Function Terms of Reference](#) (ToR). While not the primary focus of this Note, the ideas and approaches discussed here are relevant to other forms of evaluation and indeed seek to advance the consistency and quality of evaluations of scaling across CGIAR.

Box 2. Key definitions of process and performance evaluations vs. impact assessment in CGIAR

Evaluation is a systematic and objective assessment of the design, implementation and results of an ongoing or completed project, program, Initiative, policy, or operational modality in CGIAR (CGIAR Evaluation Policy, 2022).

Impact assessment research includes: (1) studies documenting the 'reach' of system level CGIAR-related innovations and policy influences in CGIAR priority countries; and (2) causal impact studies including accountability studies testing the long-term, large scale causal impacts of CGIAR research on development outcomes, and learning studies to causally test key assumptions in the theory of change of CGIAR research (The Standing Panel on Impact Assessment, SPIA [ToR](#)).

In CGIAR, process and performance evaluations do not set out to measure the impacts of scaling. Rather, these evaluations support dynamic adaptive management by providing timely evidence and advice to inform decisions as scaling unfolds. The CGIAR Evaluation Policy provides detailed definitions of process and performance evaluations.⁷ These evaluations may focus exclusively on scaling efforts (e.g., an evaluation of the S4I Program with the primary goal of coordinating and facilitating scaling of CGIAR innovations), or they may treat scaling as one component of an evaluation of a Program or Accelerator. Furthermore, scaling may be evaluated at several levels: an individual innovation, bundle of innovations, an entire Program or a part of a Program (see section 4.1.2).

As part of the development of this Note, a desk review of all 13 Programs and Accelerators in the 2025–30 Portfolio was undertaken to understand the scaling ambitions across CGIAR. All Programs and Accelerators have a scaling ambition as part of their pathway to impact alongside the five CGIAR [impact areas](#). There is a variety of scaling strategies, pathways and ambitions across CGIAR, some of which are articulated and

⁶ For example, an innovative vaccine provides individual immunity, whereas scaling vaccination produces community immunity. Evaluation is therefore needed before scaling to assess readiness, assumptions, and potential system effects such as community immunity; during scaling to monitor implementation, adaptation, and emerging risks; and after scaling to assess outcomes, sustainability, and unintended consequences. For these reasons, evaluation should span the full scaling lifecycle.

⁷ CGIAR Evaluation Policy defines process and performance evaluations as follows: Process evaluations are evaluations of the organizational functioning, instruments, mechanisms, and management practices of institutional and procedural issues across CGIAR and assessments of experience with CGIAR frameworks, policies, criteria, and procedures (CGIAR Evaluation Policy, 2022). Performance evaluations provide neutral assessments of organizational effectiveness and operating models by assessing progress toward the achievement of outcomes or processes by comparing performance data with the stated objective and reporting back on a predetermined schedule, to inform decision-making about how to best to use or invest financial or technical resources, resolve challenges, and support ongoing progress (CGIAR Evaluation Policy, 2022).

some implied (See Box 1). This demonstrates that the aims and approach of a scaling effort should guide evaluation design.

2 Clarifying Scaling Terms and Concepts

The terms and concepts presented below draw on definitions from the Monitoring, Evaluation, Learning and Impact Assessment (MELIA) [Glossary v4.1](#) (2025) and the [CGIAR 2030 Research and Innovation Strategy](#). Authors adopt broader versions of these definitions to accommodate the diversity of CGIAR's scaling efforts and to build on and connect with evaluation literature on scaling, including contributions from the Note's authors.⁸ Additional sources, such as the S4I Inception Report (Scaling Brief #4: Scaling Glossary),⁹ are also consulted.

Below are selected definitions of key terms, including 'innovation', 'scaling', 'scaling strategy', and 'scaling pathways' to keep the Note brief, additional details on these and other concepts such as 'impact', 'levels and types of scaling', and 'scaling and Innovation Portfolio Management' are provided in [Annex 2](#).

2.1 What is an Innovation?

The [MELIA Glossary](#) (2021) defines innovation as new, improved or adapted outputs or groups of outputs such as products, technologies, services, organizational and institutional arrangements with high potential to contribute to positive impacts when used at scale. Innovations can be of technological or non-technological nature and can include: (1) varieties/breeds or groups of varieties/breeds; (2) crop or animal management practices; (3) digital extension/decision support tools; (4) partnership or business models; (5) policies or other types of institutional arrangements; (6) subsidy or certification schemes; (7) capacity development programs; (8) disease detection/early warning systems; and (9) pro-poor credit arrangements, among others.

The marginal benefit of an innovation may be small or large, although some seek innovations that provide '10X', 'disruptive' or 'transformational' benefits.¹⁰ An innovation may be new or longstanding. If an innovation currently in use proves inferior to a new one, it may be necessary to down-scale the incumbent innovation to make room for the new one. However, past innovations may linger and compete with new innovations, and in some cases, they may remain superior to a newer innovation in some contexts.¹¹ Innovations are not limited to what is being scaled. The way in which scaling is undertaken may constitute an innovation. For example, the way innovations are bundled together and packaged with enabling services may also

⁸ See the International Development Research Center (IDRC) [Scaling Impact: Innovation for the Public Good](#).

⁹ Internal document not published.

¹⁰ A consultation of CGIAR Results Dashboard (accessed 3 October 2025) shows that the reported CGIAR innovation focuses on climate, environment, frontier technology, gender equality and social inclusion and market MegaTrends. Roughly 60% of innovations in 2024 were classified as technological innovation, while 22% were capacity development and 16% were policy, organizational and institutional innovations. More than two thirds (69%) of innovations were classified as incremental innovations and 7% were classified as disruptive.

¹¹ Reviewers of this Note also highlighted the need to further develop the concepts of downscaling, competition between innovations, and the superiority of non-CGIAR innovations, while keeping the Note concise. These elements could be added in a future version, after piloting, if they prove critical for evaluators.

constitute an innovation.¹² Evaluators should first have a clear understanding of what is being scaled (see sections 4.1.2 and 5.12).

2.2 What is Scaling?

In the [MELIA Glossary](#) (2021),¹³ scaling is defined as investments, strategies and processes aimed at increasing innovation readiness and/or innovation use¹⁴ to contribute to positive impacts at scale. The [Scaling Brief #4: Scaling Glossary](#) defines scaling as “a deliberate and planned effort to enable the use of innovations to have positive impact for many people across broad geographies.”¹⁵ This Scaling Brief was developed by CGIAR science leaders and scaling specialists to provide guidance for the design and implementation of major CGIAR Initiatives.

Authors invite evaluators to embrace a broader view of scaling to accommodate CGIAR’s diverse work and focus on desired changes that are sought through a scaling effort. The traditional view equates scale with growth, amplification, expansion, reach, and diffusion. This view holds that bigger is better and that an increase in the size of a Program results in a commensurate increase in its impact. The traditional view is not wrong, but it is incomplete and does not consider potential unintended negative implications of scaling.¹⁶

2.3 What is a Scaling Strategy and What are Scaling Pathways?

Scaling presents scaling actors with choices. The first choice is whether to scale the impact of an innovation. If scaling is justified, the next choice is how to scale. A scaling strategy is how scaling actors change operational scale to change impacts, and they may implement strategies alone or in combination. Some scaling strategies are in line with the traditional view because they depend on growth, reach, and diffusion. The most common of these are scaling up (increasing the number of adopters) and scaling out (operating at more sites). However, there are other strategies that depend on qualitative changes to operational scale. Scaling deep improves the quality, efficiency, or effectiveness of a Program. Transformational scaling may focus intensively on a few high leverage points in a system.¹⁷ Scaling along the causal chain may address prevention, intervention, and mitigation in one community.

Scaling pathways are the real-world mechanisms through which scaling strategies are implemented. For example, a Program could scale up by distributing an innovation to farmers through extension services or the private sector, embedding the innovation in national policies, or openly sharing the intellectual property behind the innovation with the public. Process and performance evaluators should identify intended and actual pathways, and how they may promote or hinder scaling strategies.

¹² For example, coupling technological advances with sociocultural and policy changes ([Barret et al., 2020](#)).

¹³ The 2025 MELIA Glossary v4.1 does not define scaling.

¹⁴ See Annex 1 for the definitions of Innovations readiness and innovation use.

¹⁵ [Scaling Brief #4: Scaling Glossary](#). Link accessed on 15 October 2025.

¹⁶ To explore this broad view further, see McLean, R., & Gargani, J. (2019), available for free at <https://idrc-crdi.ca/en/scaling-science> along with other scaling resources.

¹⁷ Reviewers of this Note also pointed out that transformation means fundamentally changing structures or courses of action, and thus transformational scaling would support identifying viable leverage points, as well as acting on them to shift harmful processes.

Evaluators may encounter different definitions of these concepts. They should clarify them with the evaluand team and agree early, from the inception phase, on the terms and concepts to use.¹⁸

3 Process and Performance Evaluation of Scaling: A New Area of Practice

Evaluators routinely conduct evaluations of Programs, but the evaluation of scaling efforts is new and different, so established evaluation approaches may or may not produce useful results. Here, the Note describes what makes evaluations of scaling different, starting by reviewing what is established—the underlying logic of performance evaluations applied to Programs—and then considering how it differs when applied to scaling efforts.

3.1 Familiar Ground: Evaluating Programs

Process and performance evaluations help CGIAR meet its commitments to accountability, learning, and evidence-based management, and the [IAES Evaluation Function](#) conducts these evaluations.¹⁹ These evaluations do not measure impact. Rather, they help managers judge the likelihood that a Program will subsequently achieve desirable impacts without causing undue harm, and what may be done to increase that likelihood. These evaluations follow an accepted logic that evaluators can implement in five steps (see Table 1).

3.1.1 Establish Expectations about the Program's Implementation and Results

Before Programs are implemented, researchers describe their intended actions, with whom they will take them, and the chain of consequences they expect will follow. These are their expectations. Programs may express their expectations in different ways, including logic models, theories of change (ToCs), objectives, and evaluative rubrics. These may be found in funding proposals, management reports, and other documents. Evaluators may also interview managers, scientists, and implementers to gain a more current and detailed understanding of their expectations.

3.1.2 Select Criteria that may be Used to Judge Implementation and Outcomes

With a Program's expectations in hand, evaluators work with stakeholders to select criteria that reflect the factors that matter when judging process and performance (see section 5.2). Criteria may be prioritized by different stakeholder groups to reflect what matters to them. Related evaluation questions may vary by context to reflect how process and performance are expected to vary across sites (see Table 4).

¹⁸ Evaluators may find different definitions of these concepts, for example the [Scaling Readiness Guidelines](#) defines these terms as follows: A scaling strategy is a set of coherent activities, stakeholders, and stakeholder engagement models to enable scaling. Scaling out refers to the most common way of attempting to get to scale with an innovation: reaching greater numbers of people by replication and dissemination. Scaling up refers to the attempt to change institutions at the level of policy, rules, and laws. Scaling deep refers to changing relationships, cultural values, and beliefs, addressing normative barriers to inclusive replication or scaling.

¹⁹ In February 2022, the CGIAR System Board (23rd Session) and System Council (15th Meeting) approved the new CGIAR Evaluation Framework and revised Evaluation Policy to meet CGIAR's evolving internal needs and support an organizational culture that values using performance and process evaluations to advance the public good.

Standards and principles are linked to evaluation criteria and guide evaluation practice. CGIAR developed evaluation standards and criteria, which are described in the [CGIAR Evaluation Framework and Policy](#).²⁰ These standards and principles provide a reference for professional research-for-development (R4D) evaluation and guide how evaluations are conducted within CGIAR.

3.1.3 Gather Evidence about Implementation and Results

Evaluators gather evidence about criteria performance. This may take the form of answers to descriptive questions that are based on interviews, documents, and other data. Answers are descriptive because they tell us what happened, and whether it happened well. Answers can also be quantitative, e.g., bibliometric analysis.

3.1.4 Compare the Evidence to the Standards to Reach Evaluative Conclusions

Evaluators reach systematic evaluative conclusions by comparing evidence to the standards. Comparisons may be made individually and in aggregate. In the latter case, another standard is needed to judge overall performance.

3.2 New Territory: The Logic of Process and Performance Evaluations of Scaling

Evaluations of Programs that do not involve scaling and those that do involve scaling can differ.²¹ Both apply the general logic—evidence about expected actions and their consequences is compared to standards to reach evaluative conclusions. However, evaluators must work with different kinds of expectations, criteria, standards, evidence, and comparisons that are more appropriate for scaling (see Table 1).

Table 2. How applying the logic of process and performance evaluation differs for scaling vs. non scaling programs

Five Steps	Applied to non-scaling programs	Applied to scaling programs
1. Establish expectations and assess evaluability	...about implementation and likely future results.	...about how implementation and likely future results are different than they would have been had scaling not been undertaken or undertaken in a different way.
2. Select criteria		
3. Define standards		
4. Gather evidence		
5. Compare evidence to standards		

Evaluators first establish the intended implementation and results. By contrast, to evaluate scaling, they establish how intended implementation and outcomes are expected to be different than they would have been, had scaling actors not scaled or scaled in an alternative way. Evaluators need a point of comparison

²⁰ These are also available on the [Evaluating Quality of Science for Sustainable Development Portal](#).

²¹ All CGIAR Science Programs (2025–30 Portfolio) have scaling output targets. The focus of this Technical Note extends beyond Science Programs, Future evaluations may cover other corporate, thematic, or evaluand types.

to identify what is different. The way they apply the logic of evaluation depends on their choice of comparator, and their choice of comparator depends on the evaluation questions.²²

3.3 Learning From Experience and Attending to Power

For effective scaling, it is important to build on prior experience, and allow for failures to guide anticipatory and responsive decision-making. Process evaluations can support identifying the weaknesses and opportunities in design, legitimacy and partnerships that may hinder intended adoption and impact.

Furthermore, in some cases, scaling is undertaken to create transformational system change. This suggests that those scaling innovations have great power. Evaluators have a duty to help scaling actors and use their power responsibly. Considerations of power and understanding of how power is distributed in making choices in scaling should be an important aspect of the evaluation of scaling. Who makes choices when selecting what to scale and how those choices are made are essential to understanding and evaluating. One can gain a better understanding of this duty from what historians and sociologists refer to as ‘revolutions’. The Note briefly describes two that are relevant to the CGIAR context—the Green Revolution and the Antimicrobial Revolution—and their lessons for evaluators. These examples show gaps in attention to responsible innovation during scaling. Evaluations should therefore apply a responsible innovation lens (Stilgoe et al., 2013).

Box 3. The Green Revolution

The collection of programs known as the Green Revolution are lauded for preventing potentially devastating food shortages around the world. According to Harwood (2020), the Green Revolution significantly increased crop yields, total production, and food-per-capita by the 1960s. At the same time, it promoted social inequality because of how benefits were distributed and environmental degradation because of its reliance on pesticides, herbicides, and fertilizer. Pimentel & Pimentel (1990) highlighted criticisms of CGIAR for paying insufficient attention to negative environmental impacts. Some negative impacts were anticipated from the start, and more were subsequently identified in journals and by the media. Researchers and implementers of the Green Revolution programs were slow to acknowledge these problems and were unprepared to address them (Harwood, 2020). Consequently, food systems were transformed for the better in many ways, but the Green Revolution was seen by many as reinforcing unjust social and environmental practices. With this in mind, evaluators should assess scaling efforts with the understanding that: (1) scaling creates benefits and harm for different groups in different ways in different contexts; (2) scaling actors may be overly optimistic about potential benefits and harm; and (3) part of the evaluator’s job is to help scaling actors scale responsibly by anticipating and planning for potential benefits and harm.

²² Non-scaling programs would define standards without succinctly describing scaling pathways to impact, while programs aspiring to scale would define their standards by concretely describing scaling ‘pathways to impact’.

Box 4. The Antimicrobial Revolution

Antimicrobials transformed the treatment of infectious disease in humans and animals. According to Lees et al. (2020), livestock were first treated with antimicrobial drugs in the 1930s, and the variety of antimicrobials and number of animals treated grew steadily over the next few decades. By the 1960s it was clear that scale itself was becoming a problem because the widespread use of antimicrobials created selection pressures favoring drug-resistant microbes. This is a scaling effect—a consequence of scaling an innovation, not the innovation itself. In response, microbial use was scaled back (scaled down) and protocols for proper use were tightened (scaled deep). This illustrates the dynamic nature of scaling—changes in how widely and how intensively the innovation was used changed the impacts (more animals were protected from disease), which in turn changed the context (only resistant microbes survived), which in turn affected how changing operational scale changed impacts (as reach increased, positive impacts decreased and negative impacts increased). The lesson for evaluators is to recognize that: (1) dynamic relationships like this are common; (2) they can affect impacts for better or worse; and (3) evaluators can help scaling actors anticipate, plan, and confirm them.

A [Technical Note](#) published by the Standing Panel on Impact Investment (SPIA) emphasizes that whereas the reach of an innovation in a given population may reveal something about its impact, the conditions under which reach may be a useful proxy for impact often deviate sharply from on-the-ground realities. In most settings, understanding the total benefits of an innovation requires reliable evidence of both reach and impact.²³

Thinking more broadly about scaling will help evaluators produce work that is credible, reliable and useful. Evaluators can use evidence about how inputs, processes, actions, and outputs differ across contexts to judge whether impacts will be optimized.²⁴ The broader view (see section 2.2) is preferred because it helps evaluators approach scaling more holistically. Due to the under-developed science of scaling, a lot is unknown about specific frameworks and techniques that serve the broader view. CGIAR evaluations can help scaling stakeholders understand them better and use what is learned to advance the science of scaling (Schut et al., 2020).²⁵

4 Scaling in the CGIAR Context and Implications for Evaluations

CGIAR undertakes scaling through its [2025–30 Portfolio](#). Science Programs approach scaling in different ways that take advantage of their research capabilities and alignment to Impact Areas. Accelerators collaborate with Science Programs to improve impacts through their areas of specialization—gender equity and inclusion, digital transformation, and capacity sharing. S4I helps Science Programs bundle complementary innovations, package them with supporting components, and ensure they are ready to scale in partnership with other organizations. CGIAR uses this portfolio structure to position itself as a

²³ The SPIA Note discusses the relationship between reach and impact: Does an Innovation’s reach reveal anything about its impact? Under the right conditions: possibly.

²⁴ In CGIAR’s context of R4D, while evaluating quality of science, evidence is collected on design, input, process, and outputs (see [Evaluating Quality of Science for Sustainable Development Portal](#)).

²⁵ Science of scaling is an emerging science field that CGIAR and partners are investing in.

catalyst for scaling—helping stakeholders, partners, and funders scale effective innovations collaboratively. CGIAR also directly engages in scaling efforts, such as those related to bilateral programs in which CGIAR operates part of a Program on behalf of a donor.

The following sub-sections offer a conceptual framework to assess and evaluate scaling efforts in CGIAR.

4.1.1 **Scaling Ambition**

Scaling efforts are undertaken to achieve a purpose referred to as a scaling ambition, which provides a snapshot of what is being scaled, for whom, when, with what resources, for what purpose, and in avoidance of what potential harm. It may be expressed in many ways, and not every element is relevant for every purpose. An evaluator may not find this information ready to use and may work with CGIAR stakeholders to document their scaling ambition. Evaluators should not assess results only against baseline, but also against the target(s) implied in the ambition of scale. Developing a summary statement structured in the following manner provides one approach.²⁶

Box 4. Suggested structure for scaling ambition statements

“Our ambition is to scale [the innovation] to [number, proportion, etc. of people, farms, geographies] by [target date or time span] using [resources] to accomplish [intended impacts] while avoiding [potential harm].”

4.1.2 **Innovation, Bundles, and Packages**

CGIAR develops and promotes core, complementary and enabling innovations for food, land and water system transformation.²⁷ A core innovation, such as a seed variety, technology, service, policy, or method, sits at the center of a scaling ambition. It has the capacity to produce desired impacts while avoiding harm. On its own, its capacity may be limited. Complementary innovations, such as fertilizers, markets and distribution systems, financial instruments, and algorithms, can enable, increase, or enhance impacts.²⁸

Complementary innovations are innovations used in a supporting manner. A combination of core and complementary innovations is called a bundle. Enablers are conditions, activities, or other components that support or enable successful scaling and may not be innovative. Enablers are often included in the process of innovation packaging. For example, an enabler may be a capacity sharing network or learning hub, an Accelerator to mainstream gender equality, a routine social media campaign, or a novel means of financing.

CGIAR’s scaling approach is to combine core and complementary innovations, and enablers, into Innovation Packages. The composition of an Innovation Package depends, among other things, on their readiness, market demand, and context in which scaling takes place. Decisions about what and how to scale may be informed by measures of readiness and other associated metrics, and evaluators may include them as evaluative criteria. A measure known as [Innovation Packaging and Scaling Readiness](#)

²⁶ Developing a scaling ambition is also part of what the [Innovation Packaging and Scaling Readiness](#) process offers to CGIAR and partners who engage in designing and assessing their Innovation Package.

²⁷ The [CGIAR Results Dashboard](#) (Accessed on 3 October 2025) provides detailed information about the CGIAR 2030 Program Portfolio, innovations, partnerships and results.

²⁸ When supporting innovations enhance impact, they are sometimes referred to as supplementary innovations.

(IPSR) plays a central role in CGIAR scaling metrics²⁹ (Sartas et al., 2020). Innovation bundles and packages change over time, reflecting a dynamic framework that can help teams to track scaling progress and re-prioritize scaling bottlenecks for action. The scores and their interpretations as well as a real-world example are described in Annex 3.

4.1.3 Choosing What to Scale with Evaluative Rubrics

The CGIAR scaling effort includes choices on which innovations or packages to scale or whether to advance a scaling effort to the next stage of a scaling process. These choices are made at innovation, Program or Portfolio level.³⁰ Choices are evaluative tasks, and an evaluative rubric can support them. An evaluative rubric is composed of criteria that describe characteristics of packages that make them suitable for scaling and standards that must be met for each criterion. CGIAR criteria include scaling readiness, partnership arrangements³¹, the likelihood that gender and social inclusion will be promoted, the urgency of the scaling problem, and level of risk associated with scaling. The selection depends on a variety of factors, including the availability of funds, context and purpose of scaling, co-designed with stakeholders and decisions shared with them.³² Evaluators should understand and help strengthen how these decisions are made using information available on the [CGIAR Results Dashboard](#).

4.1.4 Scaling Thesis

A scaling thesis summarizes why scaling an innovation or package will be effective, and evaluators may help CGIAR stakeholders articulate or assess their scaling thesis. The thesis has four parts: (1) what is being scaled (the innovation); (2) the pathway(s) (the real-world channels through which scaling takes place like policy or markets); (3) the scaling strategies (scaling up, out, deep); and (4) the logic or mechanisms that transform scaling into impacts (impacts add up one by one, interact in reinforcing ways to create synergies, grow in nonlinear ways like herd immunity). A scaling thesis may be structured in the following way (Box 5):

Box 5. One way a scaling thesis may be structured

“Scaling [the innovation] through [pathways] by [scaling strategies] will create [intended impacts] and avoid undue harm because [logic/mechanism].”

Hypothetical example of a scaling thesis statement: Scaling an innovative training program for farmers through a new policy by scaling out to 20 more university-based agriculture programs while scaling

²⁹ Metrics are based on [NASA’s Technology Readiness Levels](#) and was first developed for the Roots, Tubers and Bananas (RTB) Research Program.

³⁰ In CGIAR, Programs can be considered as portfolios of innovations. Portfolios can be thematic, within a CGIAR-center, country or Program.

³¹ A reviewer of this Note from the Scaling CoP emphasized that partnership arrangements should place timely emphasis on preparing for the hand-off of an innovation or intervention along the scaling pathway. In the same line, intermediaries for scaling, such as local counterpart research organizations, could play an important role if properly supported and subsequently evaluated.

³² One way to choose among many options is to score them with a rubric. In this case, there is a simple rubric with four criteria, each can earn a scores of 1 (weak), 2 (acceptable), and 3 (strong). A rubric may have a different number of categories and different scores. Standards are used to assign scores to packages on each criterion, which aggregate to an overall package score and rank. Packages with the highest scores are judged to be better candidates for scaling (see Annex 3).

deep to improve training will create good and prevent harm,³³ because strategically placing graduates in leadership roles in their communities will fundamentally change how farmers share capacity.

A scaling effort may have one scaling thesis or several rival theses that can be evaluated against each other. Every stage of a scaling process may have its own scaling thesis.

4.1.5 Scaling ToC

A scaling ToC expands the scaling thesis. It describes how scaling is expected to make impacts different than they would have been, had scaling not been undertaken or undertaken in a different way. Evaluators can help scaling actors develop, update, and test their scaling ToC, but it should not be structured exactly like a ToC for an innovation or Program, and instead should build on it. In CGIAR, scaling actors may choose either a linear or non-linear scaling ToC, and they should be able to justify their choice.

A traditional ToC maps how actions lead to outputs, outcomes, and impacts within a specific context. A linear scaling ToC shows how innovations behave differently as they scale, affecting results. While this ‘bigger-is-better’ approach may suffice in some cases, it is often limited because it assumes the relationship between scaling and context is static, overlooking their dynamic interaction (see Annex 4).

A non-linear and dynamic scaling ToC considers how scaling and the context system interact. First, if scaling is successful, outputs, outcomes, and impacts may change the context. For example, an innovative fertilizer that increases cassava yields may increase supply so much that prices fall, changing the market context. Second, scaling itself may change the context. For example, funders may consider the new fertilizer such a good social investment that it crowds out funding for other worthy innovations in the region. Third, as context changes, it has the potential to change every part of the causal process, making it more or less likely that positive impacts will be realized. For example, falling cassava prices would make it less likely that farmers will plant cassava and more likely they will switch back to their previous crops that do not benefit from the fertilizer, making it less likely that the environmental benefits of the new fertilizer will be realized. This interaction of scaling and context is what makes this scaling ToC dynamic and non-linear (see Annex 4).

4.1.6 Scaling Effects

Scaling actors often expect impacts to grow as they scale innovations. This is one type of scaling effect—impacts increasing with scale. However, scaling actors believe that impacts can grow in a variety of ways, and this has implications for evaluation. Criteria and standards used to evaluate scaling depend on the scaling effects that scaling actors expect.

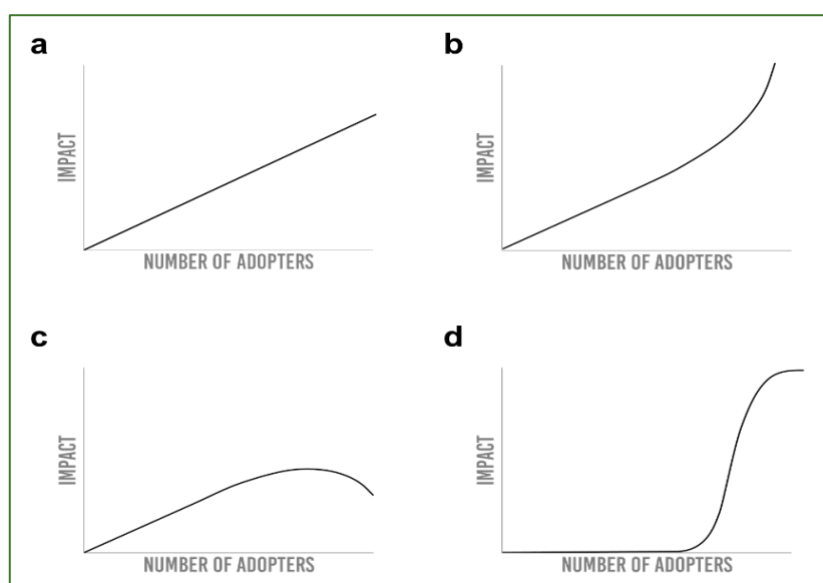
Imagine scaling actors are scaling up the number of farmers adopting a new method of water management designed to reduce water use and improve natural resource conservation. Innovators and investors may believe the best model is linear growth—for every new adopter, the total positive impact on the local environment increases by the same magnitude (4a). Alternatively, they may believe that total impact will accelerate because small farms that use less water (but have difficulty accessing it) will adopt first and larger farms that use more water will adopt later (4b). On the other hand, they may believe that

³³ Preventing harm is important; however, reviewers from CGIAR Programs also emphasize that trade-offs are inevitable at different scales and during the scaling process.

impacts will reach a maximum, then decrease if too many farmers try to independently manage the same shared resource (4c). Finally, they may believe that little will change until a tipping point is reached (4d).³⁴

The criteria and standards applied to this scaling effort would depend on the expectations that scaling actors hold. Different types of innovations may lend themselves to different expectations. CGIAR anchors its scaling efforts on the diffusion-of-innovation theory and historical adoption evidence, which posit that innovations diffuse along an S-curve and outputs follow a bell-shaped trajectory (akin to 4c).³⁵

Figure 1. Scaling dynamics: how impact changes as adoption grows



Source: adapted by John Gargani

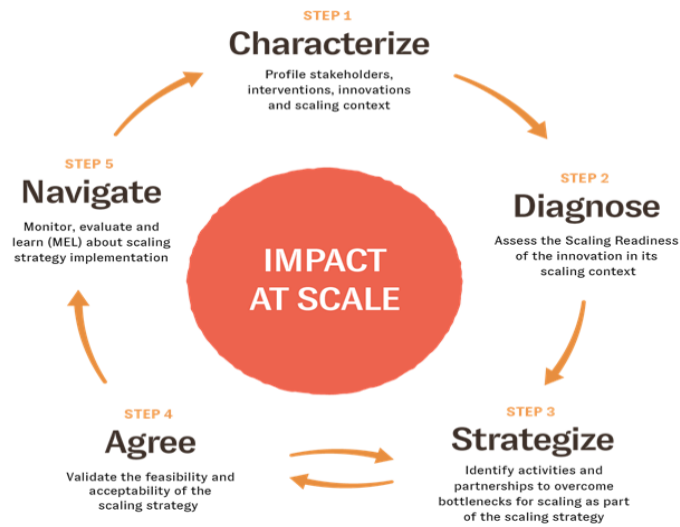
4.1.7 Mapping the Scaling Process

The scaling process may be divided into stages in different ways to capture different aspects of scaling. The five-step iterative CGIAR Scaling Process shown in Figure 5 reflects how innovation partners learn over time. Evaluators may be called upon to support continuous learning using this model.

³⁴ In most of the scaling literature, C and D are most relevant (with D being predominant with disadoption added in somewhere after the curve asymptotes).

³⁵ Reviewers of this Note indicated that this approach is currently evolving in the context of system transformation efforts.

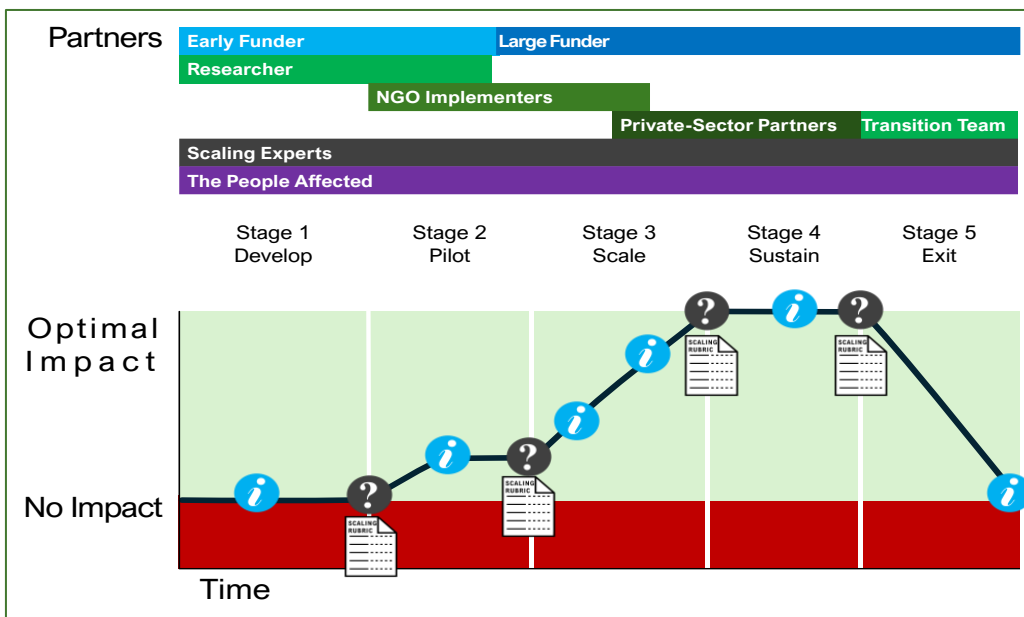
Figure 2. Scaling process



Source: *Sartas et al. (2020)*

Alternatively, evaluators may assess scaling as a sequential process that progresses through defined ‘stage gates’. In this approach, evaluators may work with stakeholders to represent stages visually, as illustrated in Figure 6. In the example shown, five stages are identified—develop, pilot, scale, sustain, and exit—although both the number and naming of stages vary in practice. Stage-gating introduces explicit decision points, represented by question marks in Figure 6, which can be supported through evaluative rubrics. Applying a rubric requires evidence, creating corresponding information points shown as ‘i’ icons in the figure. Evaluators can support stage gating by helping to develop rubrics and collecting the evidence they require.

Figure 3. Mapping the scaling process



Source: *John Gargani*

Innovation scaling requires partners, and they too can be mapped on the scaling process. CGIAR's 2025-2030 Portfolio aims to implement a more structured approach to partnerships, focusing on co-creation, enhancing capacity sharing for development, and strengthening collaboration with the private sector.

This is in line with the CGIAR 2030 Research and Innovation Strategy, and guided by the 2024 CGIAR Engagement Framework for Partnerships and Advocacy, and the 2022 CGIAR Integration Framework Agreement (IFA).³⁶ The CGIAR MELIA Glossary (2025) defines an innovation partner as organizations or entities that CGIAR collaborates and co-invests with to develop, improve the readiness of, or apply innovations to contribute to impact at scale.

Evaluation questions on partnerships could examine the extent to which partners are systematically included in Program design, implementation, and scaling. They could also explore partner roles, responsibilities, and interests, as well as the efficiency, fairness, and transparency of engagement. Additional questions might assess whether partnership processes and leadership were consistent over time and whether adequate internal capacity existed to develop and manage partnerships.

The purpose of CGIAR scaling efforts is to move innovations or packages through the stages of a scaling process, referred to as the innovation pipeline. In this case, the scaling process may be imagined as a 'leaky pipeline'—all the innovations in a batch enter the pipeline at the same stage, but only some complete all stages; the remainder exit the pipeline because they lack sufficient evidence to warrant continued scaling. The Innovation Portfolio Management approach aligns external investments in CGIAR's innovation and scaling efforts with both its short- and long-term strategies. It is a complex interplay of methods, mechanisms and mindset changes, requiring criteria and resource allocation strategies that CGIAR is currently developing but are not yet fully operational.³⁷

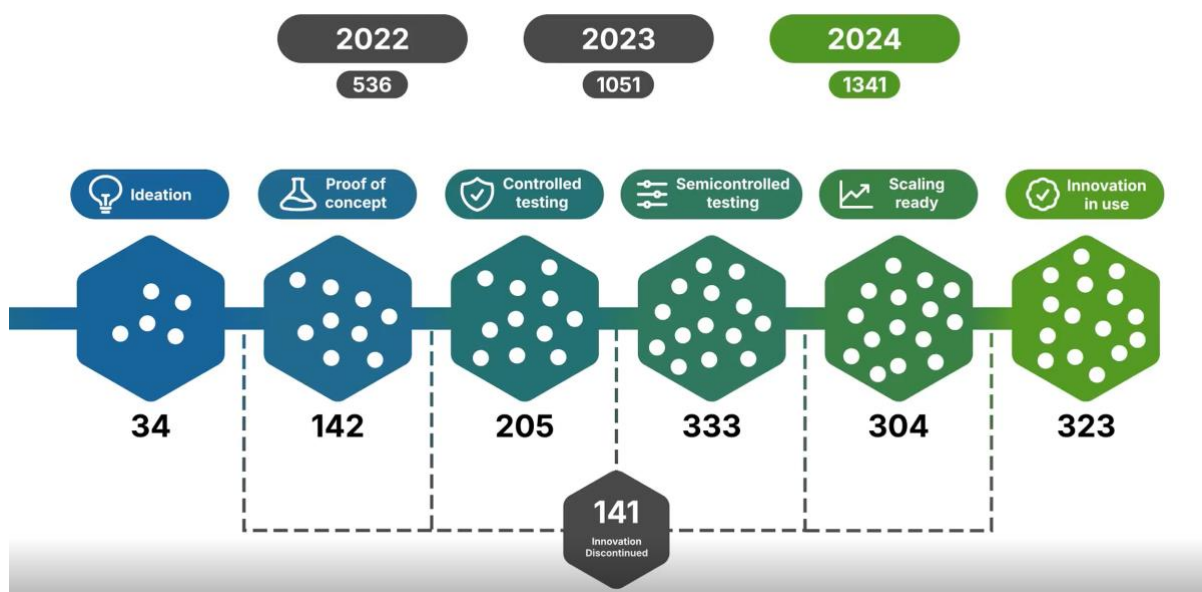
Evaluative rubrics can play two roles: (1) to guide decisions about which innovations to advance from one stage to the next; and (2) to evaluate the success of the scaling process. Metrics presented in Figure 7 may serve as criteria in the second case.

Figure 7 presents a snapshot of how innovations were distributed across stages of the innovation development and innovation use as part of 2024 W1/2 CGIAR reporting. An evaluator may identify criteria related to the figure, such as the number of innovations at each stage, their progress along levels of scaling readiness, the proportion of innovations that have been discontinued, and the characteristics that distinguish innovations at one stage from another.

³⁶ CGIAR Engagement Framework for Partnerships & Advocacy—v.2: (March 2024); CGIAR IFA v.5 (16 December 2022).

³⁷ Source: <https://www.sciencedirect.com/science/article/pii/S0308521X2400057X>

Figure 4. Snapshot of the CGIAR W1/2 innovation pipeline (2024)



Source: Marc Schut (CGIAR System Organization, shared on 6 January 2026)

The stage completion rate is the number of innovations that advance from a stage divided by the number that entered the stage. When these metrics are used as criteria, evaluators must establish standards for them. For example, stakeholders may judge a stage 1 completion rate of 90% or more to be excellent and 40% or less to be poor. Without a standard, evaluators can describe the process and performance of scaling efforts, but they cannot reach evaluative conclusions. It may be difficult, however, to find benchmarks to guide standard setting, and standards will vary by context, riskiness, and urgency.

It may take years for every innovation to either complete the pipeline or exit. Defining stages, criteria, and standards at the start of a scaling process makes it possible to produce periodic evaluations of the progress innovations made throughout a development pipeline. Reviewers of this Note emphasized the need to assess and support decision-making that advances innovations along the pipeline, even at low readiness levels, and to examine the effectiveness, legitimacy, and inclusiveness of scaling pathways. [CGIAR QR4D Framework to Process and Performance Evaluations](#) provides a good framework for integrating these aspects in the evaluation design.

5 Framework for Evaluating Scaling Efforts

5.1 How to Prepare

During the inception phase of an evaluation, evaluators should become familiar with CGIAR scaling tools and the reference documents compiled in this Note and related reference materials and inquire for any updates post release of this Note (February 2026).

The steps below are intended to help evaluators structure their learning about what is being evaluated—scaling efforts. Not all steps need to be followed; their relevance depends on what is being scaled, the

scope and type of scaling effort, and the context. The main concepts and tools are described in this Note with additional elements provided in the annexes.

Table 3. Key steps to preparing for the evaluation

#	Key question	Related concepts/tools
1	What is being scaled?	Innovations, bundles, and packages
2	How was it chosen?	Scaling readiness; evaluative rubric for scaling
3	What is the ambition?	Scaling ambition statement
4	Why will it work?	Scaling thesis; scaling ToC; scaling effects
5	How will scaling unfold?	Scaling thesis; scaling process map
6	Who are the partners?	Scaling process map
7	What is successful scaling?	Evaluative rubric for scaling
8	What is the portfolio of innovations being developed for scaling?	Leaky pipeline model; transition matrix

Evaluation managers and evaluators should assess the evaluability of scaling efforts as early as possible, during the development of the ToR and continuing through the inception stage. Where available, they may also rely on evaluability assessments already carried out by the IAES Evaluation Function³⁸ or by the evaluand team. Key elements to consider include: a clear scaling ambition and scaling thesis; a well-defined ToC; MELIA systems that track progress along scaling pathways (not only outputs); clarity on partner and stakeholder roles, including voice in scaling decisions; and adequate data, timing, and access to assess changes in operational scale and expected impacts. For further guidance, evaluators can consult CGIAR guidelines [Conducting and Using Evaluability Assessments in CGIAR](#).³⁹ Evaluators are also encouraged to review findings from the 2025–30 Portfolio evaluability assessments, which highlighted recurring challenges related to scaling ambition, nested ToCs, and MELIA systems for scaling.

5.2 Criteria, Principles and Guiding Questions

The Note organizes evaluation questions on scaling around evaluation criteria, principles, and guiding questions. Process and performance evaluations of scaling within CGIAR may address a selection of these questions. Evaluators are not expected to use all questions or criteria but should select those that best fit

³⁸ 2025 Evaluability Assessments of all Programs and Accelerators under the [CGIAR's 2025–30 Portfolio](#) under the [2030 Research and Innovation Strategy](#) were included in the System Council-approved Multi-Year Evaluation Plan (MYEP) ([2025–27 Work Plan for CGIAR's Independent Advisory and Evaluation Service \(IAES\) \(SC/M21/DP5\)](#)). Midline evaluations of the Programs and Accelerators are scheduled for 2027.

³⁹ The Framework's seven Domains include: A. Intervention Logic; B. MEL Systems and Resources; C. Gender Diversity and inclusion; D. Long-term evaluability; E. Context and Environment; F. Management and Stakeholder Engagement and Response; and G. Cost Effectiveness and Efficiency (see page 5).

the evaluation purpose, focus, and available evidence and can suggest new or adjusted questions. The mapping of questions and principles to criteria is indicative rather than strict, as some questions may relate to multiple criteria or principles. The same applies to the principles, which may also cut across different criteria.

This framework was grounded in the [CGIAR Evaluation Framework](#) and the [CGIAR Evaluation Policy \(2022\)](#), as well as the guidelines on applying the CGIAR Quality of Research-for-Development (QoR4D) Framework to process and performance evaluations. The OECD-DAC evaluation criteria (OECD, 2019) are used, with questions adapted to the CGIAR scaling context and to the concepts presented in this Note. The principles and questions were grounded in CGIAR MELIA reference documents and scaling literature. They were jointly identified by CGIAR scaling colleagues and the IAES Evaluation Function responding to known challenges and dimensions in the CGIAR context and Scaling Portfolio. Reviews of the Note also informed the identification of key scaling matters from different perspectives and backgrounds, which were translated into additional evaluation questions.

Table 4. Scaling evaluation criteria, principles and questions

Evaluation Criteria	Principles	Evaluation questions	
Relevance and Coherence	A. Demand- and user-centered design	1. To what extent has the problem been validated with key stakeholders (partners, clients, beneficiaries, end users), and how well have problem drivers and root causes been identified?	
		2. Is demand intelligence available, and how has it informed the definition and prioritization of the problem and proposed solutions?	
		3. Is there clarity on what is being scaled, including the innovation or innovation bundle, its intended function, and its target users?	
		4. Is there an agreed scaling ambition that has been validated with key demand, scaling partners, and other relevant stakeholders?	
Efficiency	B. System perspective	5. Are there sufficient capacity and investment to manage innovations and Innovation Portfolios for scaling in an evidence-based manner?	
		6. To what extent have multiple innovations been systematically assessed for their ability to address the identified problem?	
		7. To what extent has scientific and other relevant evidence been used to assess, compare, and prioritize innovations or solution options for scaling?	
		8. To what extent have innovations been bundled and adapted to local contexts to support the agreed scaling ambition or impact objectives?	
		C. Evidence-based Portfolio Management	9. Are system bottlenecks clearly identified, and how realistic are the assumptions that these bottlenecks can be addressed within available time and resources?
			10. Are there an adequate framework and system for assessing, managing, and improving innovation and scaling performance and are they effectively operationalized?
			11. Are roles and responsibilities for production, delivery, adoption support, and enabling policy clearly assigned and effectively performed by CGIAR and its partners to sustain scaling over time?

Evaluation Criteria	Principles	Evaluation questions
Effectiveness	D. Responsible scaling	12. To what extent are demand and scaling partners actively involved in co-creating the scaling pathway(s), in ways that place them in a leading role?
		13. Is there a due diligence process to identify appropriate scaling partners and assess their capacity and willingness to co-invest in durable change or impact at scale?
		14. How well are evaluand's scaling ambitions aligned with the capacities and practices of its main partners?
	E. Effective partnerships and networks	15. To what extent were partnerships building consistent and continuous over time?
		16. To what extent was there sufficient internal capacity to develop and manage partnerships?
	F. Responsible scaling to gender, diversity, and inclusion	17. Is there clarity on who is likely to benefit from scaling?
		18. To what extent are mitigation measures in place to uphold a 'do no harm' principle (e.g., for vulnerable populations or ecosystems)?
		19. To what extent are potential unintended negative consequences or trade-offs identified, monitored, and managed throughout the scaling process? To what extent are gender, diversity, and inclusion clearly defined and reflected within the evaluand's scaling ambition, ToC, and stakeholder engagement?
	G. Do no harm	20. To what extent do existing data, indicators, and MELIA systems enable the assessment of differential gender and inclusion outcomes resulting from scaling efforts?
		21. Were the necessary inputs adequate and sufficient to deliver the planned research outputs and outcomes?
Quality of Science (scaling)	H. Credible and legitimate science of scaling	22. Was the composition of research and delivery teams appropriately diverse?
		23. To what extent did scientific outputs contribute to science-based innovations, targeted capacity development, and policy advice relevant to scaling?
		24. To what extent were research findings effectively communicated and disseminated to relevant audiences?
		25. To what extent do the outputs reflect attention to gender, diversity, and inclusion and/or environmental considerations?
		26. How do limitations in mandate, resources, and capacity affect the achievement of evaluand's scaling ambitions?
Impact	I. Strategic and diversified scaling pathways	27. To what extent have multiple scaling pathways been identified and assessed for their potential to generate durable impact?
		28. Is credible financial and market evidence (e.g., market sizing, cost-benefit analysis, willingness to pay) available to support the scaling investment plan?

Evaluation Criteria	Principles	Evaluation questions
Sustainability	J. Sustainable and achievable solutions	29. Is there a clear and credible vision or strategy for achieving long-term or durable impact at scale without continued direct involvement of CGIAR? 30. Is there a realistic exit strategy for CGIAR and its core partners, supported by clear commitments and financial plans from scaling partners to sustain scale beyond project or Program funding?

Process and performance evaluations help CGIAR meet its commitments to accountability, learning, and evidence-based management. Rooted in the CGIAR Evaluation Framework and Policy (2022), the Note describes what makes evaluations of scaling different, starting by reviewing what is established—the underlying logic of performance evaluations applied to Programs—and then considering how it differs when applied to scaling in the context of CGIAR. The Note sets out the main approaches. However, the evaluation of scaling is still at an early stage, as the science behind it is still developing, and such evaluations remain rare. Ongoing learning from practice and from emerging standards will be important to improve how scaling is evaluated. IAES will update the Note as CGIAR gains experience from its use and as new developments in scaling practice emerge.

References

- Antle, J. M., & Valdivia, R. O. (2021). Trade-off analysis of agri-food systems for sustainable research and development. *Q Open*, 1(1), qaaa005. <https://doi.org/10.1093/qopen/qaaa005>
- Barrett, C. B., Benton, T. G., Cooper, K. A., Fanzo, J., Gandhi, R., Herrero, M., James, S., Kahn, M., Mason-D'Croz, D., Mathys, A., Nelson, R. J., Shen, J., Thornton, P., Bageant, E., Fan, S., Mude, A. G., Sibanda, L. M., & Wood, S. (2020). Bundling innovations to transform agri-food systems. *Nature Sustainability*, 3(12), 974–976. <https://doi.org/10.1038/s41893-020-00661-8>
- Brooks, M., Lybbert, T.J., & Walker, M. (2025). Does an Innovation's Reach Reveal Anything About its Impact? Under the Right Conditions: Possibly. Technical Note N.12. Rome. SPIA. <https://cgspace.cgiar.org/server/api/core/bitstreams/51db0e95-74f5-415a-bf34-3267b72feebd/content>
- CGIAR. (2024). Toward Greater Collective Impact: A CGIAR Engagement Framework for Partnerships & Advocacy—Version 2.0: March 2024. Available at: [CGIAR Resource](#) (Accessed 2 October 2025).
- CGIAR. (2023). CGIAR Integration Framework Agreement (IFA) v.5 – 16 December 2022. Available at: [CGIAR Resource](#) (Accessed 2 October 2025).
- CGIAR Independent Advisory and Evaluation Service (IAES). (2025). Assessing the Evaluability of CGIAR's 2025–2030 Portfolio: Need Assessment Results. Rome: IAES Evaluation Function. <https://iaes.cgiar.org>
- CGIAR Independent Advisory and Evaluation Service (IAES). (2025). *Summary of Learnings on Monitoring, Evaluation, Learning and Impact Assessments (MELIA)*. Rome: IAES Evaluation Function. <https://iaes.cgiar.org/evaluation>
- CGIAR Independent Advisory and Evaluation Service (IAES). (2023). Regional Integrated Initiative—Evaluability Assessment Report. Rome: IAES Evaluation Function. <https://cas.cgiar.org/>
- Fuglie, K. O., & Echeverria, R. G. (2024). The economic impact of CGIAR-related crop technologies on agricultural productivity in developing countries, 1961–2020. *World Development*, 176, 106523. <https://doi.org/10.1016/j.worlddev.2023.106523>
- Gargani, J., & King, J. (2024). Principles and methods to advance value for money. *Evaluation*, 30(1), 50–68. <https://doi.org/10.1177/13563890231221526>
- Ghiron, I., E. Ramirez-Ferrero, R. Badiani, R. Benevides, A. Ntabona, P. Fajans, R. Simmons (2021): Promoting scale-up across a global project platform: lessons from the evidence to action project Glob. Implement. Res. Appl., 1 (2), pp. 69–76
- Harwood, J. (2020). Could the adverse consequences of the green revolution have been foreseen? How experts responded to unwelcome evidence. *Agroecology and Sustainable Food Systems*, 44(4), 509–535.
- Hearn, S., Lomofsky, D., Chames, C., Datta, A., Davies, N., & Consulting, O. T. T. (2021). *Evaluation of the International Development Research Centre's strategy to scale research results*. <http://hdl.handle.net/10625/60649>
- ISDC. 2022. Application of the Quality of Research for Development framework to Initiative reviews: Lessons learned. Rome: Independent Advisory and Evaluation Service. [IAES](#)

- Kangethe, E., & Flintan, F. (2022). *Participatory land use planning in pastoral areas in Tanzania: IPSR innovation package and scaling readiness report* (1st ed.). CGIAR System Organization.
<https://hdl.handle.net/10568/128065>
- Kohl, R. (2021). *Scaling and Systems: Issue Paper*, Global Community on Scaling Development Outcomes
- Lees, P., Pelligand, L., Giraud, E., & Toutain, P. L. (2021). A history of antimicrobial drugs in animals: Evolution and revolution. *Journal of Veterinary Pharmacology and Therapeutics*, 44(2), 137-171.
- Leeuwis, C., B.K. Boogaard, and K. Atta-Krah (2021): How food systems change (or not): governance implications for system transformation processes. *Food Security*, 13 (4), pp. 761-780
- Low, J.W. and G. Thiele (2020): Understanding innovation: the development and scaling of orange-fleshed sweet potato in major African food systems, *Agric. Syst.*, 179, 10.1016/j.agry.2019.102770
- Lynam, J., Byerlee, D., & Mook, J. L. (2024). The organizational challenge of international agricultural research: The fifty-year odyssey of the CGIAR. *Food Policy*, 124, 102617.
<https://doi.org/10.1016/j.foodpol.2024.102617>
- Marc Schut, Cees Leeuwis, Graham Thiele (2020): Science of Scaling: Understanding and guiding the scaling of innovation for societal outcomes, *Agricultural Systems*, Volume 184, 102908, ISSN 0308-521X
- McIntire, J., & Dobermann, A. (2023). The CGIAR needs a revolution. *Global Food Security*, 38, 100712.
<https://doi.org/10.1016/j.gfs.2023.100712>
- McLean, R., & Gargani, J. (2019). *Scaling impact: Innovation for the public good*. Routledge.
- Pimentel, D., & Pimentel, M. (1990). Comment: Adverse environmental consequences of the Green Revolution. *Population and Development Review*, 16, 329-332.
- Sartas, M., Schut, M., Proietti, C., Thiele, G., & Leeuwis, C. (2020). Scaling readiness: Science and practice of an approach to enhance impact of research for development. *Agricultural Systems*, 183, 102874.
<https://doi.org/10.1016/j.agry.2020.102874>
- Sartas, M., Schut, M., van Schagen, B., Velasco, C., Thiele, G., Proietti, C., & Leeuwis, C. (2020). *Scaling Readiness: Concepts, practices, and implementation*. CGIAR Research Program on Roots, Tubers and Bananas (RTB).
- Scaling Community of Practice (2022): *Scaling Principles and Lessons*
- Schut, M., Adeyemi, I., Kumpf, B., Proud, E., Dror, I., Barrett, C. B., Menzies, D., Colomer, J., & Leeuwis, C. (2024). Innovation Portfolio management for the public non-profit research and development sector: What can we learn from the private sector? *Innovation and Development*, 0(0), 1-19.
<https://doi.org/10.1080/2157930X.2024.2400779>
- Schut, M., Leeuwis, C., & Thiele, G. (2020). *Science of scaling: Understanding and guiding the scaling of innovation for societal outcomes* [Special issue]. *Journal Name*.
<https://www.sciencedirect.com/special-issue/108MNQ3QF2N>.
- Schut, M., Colomer, J., Proud, E., Bonaiuti, E., Dror, I., Kang'ethe, E., Esquivias, L., & Leeuwis, C. (2024). *Innovation Portfolio management for responsible food systems transformation in the public sector: Lessons, results and recommendations from CGIAR*. *Agricultural Systems*, 216, Article 103907.
<https://doi.org/10.1016/j.agry.2024.103907>

- Scaling Community of Practice. (2025). *Evaluation guidelines of official international development funders*. <https://scalingcommunityofpractice.com/wp-content/uploads/2025/04/FINAL-Evaluation-Guidelines-of-Official-International-Development-Funders.pdf>
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). *Developing a framework for responsible innovation*. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Woltering, L. (2024). *Stuck on scale: Rethinking scaling and systems change for Agricultural Research for Development for food systems transformation* [PhD Thesis, Wageningen University]. <https://research.wur.nl/en/publications/stuck-on-scale-rethinking-scaling-and-systems-change-for-agricult>

Annex 1. CGIAR MELIA Glossary v4.1 (September 2025)

The MELIA Glossary serves as an interim solution to ensure consistent taxonomy across CGIAR guidance documents, tools, and platforms for technical reporting, until the CGIAR taxonomy, being co-developed by centers and the System Organization under the Monitoring, Evaluation, Learning, Impact Assessment, and Foresight (MELIA-F) Project, is finalized (expected mid-2026).

Table 5. CGIAR MELIA Glossary v4.1 (September 2025) for scaling

Term	Definition
Accelerator	Accelerators will undertake strategic research in their topical areas and, through working with all other Programs, bolster CGIAR's ability to reach and support targeted end-users while furthering collaboration, coherence, and integration across the entire Portfolio.
Impact	A durable change in the condition of people and their environment brought about by a chain of events or change to which research, innovations and related activities have contributed. Impacts may be positive or negative, direct or indirect, intended or unintended. Examples include: Increased average household income in the target demographic; reduced greenhouse gas emissions from target landscapes; and increased number of women farmers with secure rights to land.
Impact Area	CGIAR is targeting multiple benefits across five Impact Areas. Working with others, for each of the Impact Areas, CGIAR contributes to collective global targets for the transformation of food, land, and water systems across local, regional, and global levels. CGIAR Impact Areas are: (1) Nutrition, Health and Food Security; (2) Poverty Reduction, Livelihoods and Jobs; (3) Gender Equality, Youth and Inclusion; (4) Climate Adaptation and Mitigation; and (5) Environmental Health and Biodiversity.
Indicator	A standardized measurement that provides information on progress or change towards a specific objective or issue. These may be used to trace and compare results, monitor trends, identify areas for action across impact pathways over time and across contexts.
Innovation	New, improved or adapted outputs or groups of outputs such as products, technologies, services, organizational and institutional arrangements with high potential to contribute to positive impacts when used at scale. Examples include: innovations can be of technological or non-technological nature and can include varieties/breeds or groups of varieties/breeds; crop or animal management practices; digital extension/decision support tools; partnership or business models; policies or other types of institutional arrangements; subsidy or certification schemes; capacity development programs; disease detection/early warning systems; and pro-poor credit arrangements.
Innovation Development	The process of developing innovations from an idea to a proven product, technology, service or arrangement that is validated for its ability to contribute to positive impacts when used at scale. The process often includes basic research, innovation design, and improving innovations based on testing under (semi-)controlled and uncontrolled conditions. Progress in innovation development is measured in an evidence-based way along the 0-9 levels of innovation readiness.
Innovation Package	Combinations of interrelated innovations and enabling conditions that, together, can lead to transformation and impact at scale in the CGIAR research delivery hierarchy. They are context, outcome, and use-group specific and their ability to contribute to outcomes and impact can change over time. Innovation Packages is a key concept of the Scaling Readiness approach .

Term	Definition
Innovation Partner	Organizations or entities that CGIAR collaborates and co-invests with to develop, improve the readiness of, or apply innovations to contribute to impact at scale.
Innovation Readiness	<p>A metric used to assess the maturity of an innovation, with a scale ranging from the idea (lowest level) to validated under uncontrolled conditions (highest level).</p> <p>Examples: Innovation Readiness is a key metric in the Scaling Readiness approach. Innovations often go through different stages of readiness and can start as an idea that is turned into a prototype, working model or application that needs to be tested and validated under controlled and uncontrolled conditions. The readiness of an innovation is often context-, and outcome specific, and can change over time. Innovation Readiness is assessed using evidence and prevents the scaling of innovations that are not proven to work.</p>
Innovation Use	<p>A metric used to assess the extent to which an innovation is already being used, by which type of users and under which conditions, with a scale ranging from no use (lowest level) to common use (highest level). Examples: innovation use is a key metric in the Scaling Readiness approach.</p> <p>An innovation that is only used by the project, design or intervention team or its direct partners will score low in innovation use. Innovations that are commonly used by anticipated end-users will score high in innovation use. Users that are directly incentivized by a project or intervention to use an innovation are considered project team or direct partners which will score as low innovation use.</p>
Innovation Packaging and Scaling Readiness (IPSR) Pathway to report Innovation Use (outcome)	<p>An advanced way to report progress in innovation use towards impact at scale using the Innovation Packages and Scaling Readiness approach. This pathway supports CGIAR and partner teams in:</p> <ul style="list-style-type: none"> • Co-creating a scaling ambition for innovations with scaling partners. • Designing a context-specific Innovation Package (enabling conditions for scaling) with partners and experts. • Assessing the Innovation Package for its scaling readiness (evidence-based). • Identify bottlenecks for scaling and partner discussions on how to overcome those. • Reporting current use of the core innovation (evidence-based).
Non-IPSR Pathway to report Innovation Use (outcome)	<p>Supports reporting the current use of a core innovation (evidence-based) but does not enable Programs/Accelerators/projects to report progress in innovation use towards impact at scale, and does not support CGIAR and partner teams in:</p> <ul style="list-style-type: none"> • Co-creating a scaling ambition for innovations with scaling partners. • Designing a context-specific Innovation Package (enabling conditions for scaling) with partners and experts. • Assessing the Innovation Package for its scaling readiness (evidence-based). <p>Identify bottlenecks for scaling and partner discussions on how to overcome them.</p>
Outcome	<p>A change in knowledge, attitudes, skills, and/or relationships, which manifests as a change in behavior within the spheres of influence and interest, that results in whole or in part from the research and its outputs. Examples include: use of a new technology (including outputs like a seed variety) by farmers; policy actors using research-based knowledge to inform policy decisions; participants in a CGIAR-supported process agree to new germplasm conservation and exchange protocols; and researchers use CGIAR generated methods and/or databases. Key outcomes include: Who will do what differently as a result of the Program/Accelerator/project?</p>

Intermediate outcome	Intermediate outcomes are necessary steps towards achieving the 2030 outcomes. Intermediate outcomes will lead to the 2030 outcomes, although the actors and units of measure may be different.
2030 Outcome	2030 outcomes are the final, highest-level outcomes to be achieved by the end of 2030 by the Programs/Accelerators/projects. It is expected that these 2030 outcomes would lead to measurable impacts.
Output	Knowledge, technical or institutional advancement produced by CGIAR research, engagement and/or capacity development activities. Examples include new research methods, policy analyses, gene maps, new crop varieties and breeds, institutional innovations or other products of research work, and partnerships because of a signed memorandum of understanding.
High-level Output	High-level outputs are aggregated clusters of outputs which are linked by the intention to achieve a specific outcome. High-level outputs should be relatively downstream—i.e., ready to be used by next/end users—and would often be composed of various types of outputs (e.g., knowledge products, innovations, and capacity sharing activities).
Partner	Organizations or individual stakeholders that CGIAR collaborates with to achieve its goals for which proof exists, i.e., memorandum of understanding, joint code of practice, a contract under a joint proposal. Partners can also be beneficiaries of CGIAR’s interventions. Examples include demand partner, innovation partner and scaling partner.
Partnership	Proven relationships between CGIAR’s entities and/or partners, to jointly undertake activities that contribute to a common set of objectives or a common goal along the lines of each partner’s mandate. A partnership can include multiple partners. Proof of partnerships should exist in the form of any signed document explaining the common set of objectives and/or goals for which the partnership exists. See also: Partners.
Performance and Results Management Framework	The Performance and Results Management Framework (PRMF) describes the processes, systems and measures for managing CGIAR’s performance and results to support delivery of the CGIAR 2030 Research and Innovation Strategy . It helps teams to monitor, learn and adapt their Science Program/Accelerator over time to best navigate towards impact, support leaders to allocate resources towards workstreams with growing potential, and help CGIAR communicate the full spectrum of its impact to partners, funders and beneficiaries.
Performance and Results Management System	The Performance and Results Management System (PRMS) underpins the PRMF, encompassing planning, monitoring, and reporting. It provides robust information to support informed decision-making, enabling real-time data collection and the day-to-day management of the Portfolio and Science Programs/Accelerators.
Policy Change	Policies, strategies, legal instruments, Programs, budgets, or investments at different scales (local to global) that have been modified in design or implementation, with evidence that the change was informed by CGIAR research.
Portfolio	The CGIAR System Framework approved by the System’s funders and centers in June 2016, and then amended in October 2024, defines the CGIAR Portfolio as the research Programs and/or platforms carried out by the centers and CGIAR System partners in support of the CGIAR Strategy and Results Framework and are supported by (1) the CGIAR Trust Fund; and/or (2) bilateral sources contractually aligned to such Programs and/or platforms.
Program	CGIAR Programs are the main vehicle for delivery of research and innovation by CGIAR from 2025–30. Programs serve as entry points to describe CGIAR’s offer on a key topic, elevating CGIAR’s visibility in global agendas and facilitating the continuation and formation of inclusive alliances and partnerships. They state quantitatively what impacts and outcomes they intend to achieve, along a theory of change (ToC). They come with evaluable results frameworks and clear reporting of results against investment. Bilateral and window 3 projects are aligned to

Programs and Accelerators, and their results are reported as part of Program and Accelerator reporting.

Result	The product of an activity by CGIAR staff or due in part to CGIAR staff, described as an output, outcome or impact.
Scaling Partner	Organizations or entities that CGIAR collaborates with to advance the uptake and use of innovations at scale.
Scaling Readiness	A metric that combines single or average innovation readiness and innovation use scores at Innovation Package or portfolio level (see also: Innovation readiness and innovation use). More information at www.scalingreadiness.org .
Sustainable Development Goal (SDG)	Seventeen (17) internationally accepted , publicly broadcast objectives related to achieving sustainable development adopted from 2015–30.
ToC	An explicit, testable model of how and why social change is expected to happen along an impact pathway in a development context. A basic research-for-development ToC identifies the context and key actors in a project, Program or initiative and specifies the causal pathways and mechanisms by which the research aims to contribute to outcomes and impacts. This model can be used for: (1) designing and planning project logic including assumptions; (2) monitoring and evaluation of processes and results including risk identification and mitigation; (3) testing assumptions and identifying lessons learned, reflection for accountability and results-based management, and using findings for evidence-based decision-making for learning and accountability.
Quality Assurance	Quality assurance (QA) is a process to monitor, validate and improve the quality of data based on established procedures and guidelines. Within CGIAR's overall technical reporting process, the QA includes: <ul style="list-style-type: none"> • Evaluating selected reported data to ensure their accuracy and alignment with the evidence presented based on agreed-upon criteria. • Providing guidance to improve the data entry process in the pre-QA phase of the Technical Reporting process. • Continuously improving CGIAR's QA process to increase its effectiveness, efficiency, and cost-effectiveness.

Annex 2. Additional Information about Terms and Concepts

What is Scaling?

Acting differently: The Note calls this changing operational scale to distinguish it from scaling impact. When programs, organizations, sets of innovations, and partnerships scale, they operate differently than they would have had they not scaled or scaled in another way.⁴⁰ Process and performance evaluators seek to understand how programs operate differently when they scale. Another distinction evaluators should pay attention to is between research, development and scaling efforts and how each might presuppose different processes and decisions.

⁴⁰ This can also be understood in comparison with programs that focus primarily on research, where scaling is not an objective.

Outcomes that are different from what they would have been had operational scale not been changed or changed in another way: Outcomes⁴¹ may differ in their magnitude (the number, size, or reach of impacts), variety (which effects are created), equity (differential benefits and harm across individuals, groups, and geographies), sustainability (the longevity of impacts and their contribution to the wellbeing of the planet), and other qualitative and quantitative dimensions.

The CGIAR Guidelines [Applying Quality of Research for Development \(QoR4D\) Framework to Process and Performance Evaluations](#) suggest **seven criteria organized around four elements** that evaluators may use to describe scaling outcomes. CGIAR process and performance evaluations seek to understand how outcomes of research-for-development (R4D) manifest as a change in behavior in actors—acting differently, and whether acting differently is likely to make outcomes different in anticipated and unanticipated ways. Evaluators also assess the scalability of an innovation as well as progress along scaling pathway(s).

Better impacts are those that **the people affected judge to be better**. Should scaling actors and the people affected hold different views, evaluators can help them find common ground and/or evaluate from multiple perspectives by using a different set of criteria and standards for each group.⁴² CGIAR process and performance evaluators may not have the time and resources to engage with the people affected by scaling, but they have a duty to assess whether it is likely they will judge anticipated differences in impacts favorably. Similarly, scaling actors undertaking demand-driven scaling, for example, should be able to provide evidence that scaling will change impacts in ways that consumers and others value.

What is Impact?

Development professionals define impact differently⁴³, and CGIAR specifically holds a broader view (MELIA Glossary, 2025).

A durable change in the condition of people and their environment brought about by a chain of events or change to which research, innovations and related activities have contributed. Impact may be positive or negative, direct or indirect, intended or unintended.

In CGIAR, the Standing Panel on Impact Assessment (SPIA) has the mandate to conduct or commission impact assessment research including: (1) studies documenting the reach of CGIAR-related innovations and policy influences at system level in CGIAR priority countries; (2) causal impact studies including accountability studies testing the long-term, large scale causal impacts of CGIAR research on development outcomes; and (3) learning studies to causally test key assumptions in the theory of change (ToC) of CGIAR research.

This Note uses the term impact in its broad sense, since it is better suited for scaling. Impact encompasses intentions that are relatively modest (such as increasing the adoption of a higher yielding crop variety on a

⁴¹ The MELIA Glossary (2021) defines outcome as: a change in knowledge, skills, attitudes and/or relationships, which manifests as a change in behavior in particular actors, to which research outputs and related activities have contributed. Examples of outcomes include: use of a new technology by farmers; policy actors using research-based knowledge to inform policy decisions; participants in a CGIAR-supported process agree to a new germplasm conservation and exchange protocols; and researchers using CGIAR-generated methods and/or databases.

⁴² For a discussion on how to apply multiple criteria and standards, see [Gargani & King \(2024\)](#). The context is valuable for value for money evaluations, but the approach may be applied to any evaluation.

⁴³ [Hearn, Simon & Buffardi, Anne. \(2016\)](#). What is impact?

small scale) to those that are highly ambitious (such as transforming regional food systems). Moreover, equity, inclusion, sustainability, and social justice fit within a broad conception of impact because they are possible consequences of development actions.⁴⁴

What are Levels and Types of Scale?

Scaling may also cause impacts to progress through levels and types. The term **level of scale** is used in several ways that emphasize different aspects of scaling. In CGIAR,⁴⁵ a level of scale may refer to domains that are nested from low to high, such as fields which are nested within farms which are nested within communities. From this perspective, a program scales by progressively operating at higher levels of scale. Level of scale may also mean simply doing more, for example a factory increasing its production from 100 bags of fertilizer a day to 500, or a gender equity program doubling the number of communities it serves. From this perspective, a program scales by progressively operating at greater levels of scale to affect more people, places, and things.

Sometimes acting differently when scaling means operating in qualitatively different ways, which some reviewers of the Note consider a form of scaling deep. For example, improving the effectiveness of training or deepening stakeholder inclusion are qualitative changes to operations that can scale impact. From this qualitative perspective, a Program scales by progressively operating at different types of scale.

For example, scaling may cause an impact to move from a lower level (individual health) to higher (community health), from a smaller level (few farmers) to greater (many farmers), and from one type (inequitable) to another (equitable). Scaling actors often scale by simultaneously operating at higher levels, greater levels, and different types of scale. Similarly, scaling may include changes of levels and types, simultaneously. This is why evaluators should examine both quantitative and qualitative changes to operational scale and expected impacts when assessing scaling efforts.

What is Scaling and Innovation Portfolio Management?

Innovation Portfolio Management refers to the active management of research, innovation and scaling investments along an organizational, Program or project vision or strategy. It covers the entire idea-to-impact-pathway, not just the scaling stages. Similar to innovation pipeline management or stage-gating, Innovation Portfolio Management assesses the scalability potential of innovations from early stages onward. The aim is to prioritize (scarce) resources, ensuring portfolio health, managing risks, and to learn from failure and success to increase innovation and scaling performance.

Most projects, Programs and organizations develop, test and scale multiple innovations at the same time. These collections of innovations are often referred to as Innovation Portfolios. Innovation Portfolios can have geographic (country), thematic (climate), programmatic (breeding) or other organizing themes.

Innovation Portfolios often include innovations across different readiness and/or use levels. Innovation Portfolios often aim to achieve a specific organizational or Program vision or strategy. Innovation portfolio management then includes making decisions on resource allocation to ensure portfolio balance/health and to ensure that innovations that demonstrate high potential for impact are prioritized and adequately resourced, whilst deprioritizing or stopping innovations that do not work or show less potential.

⁴⁴ Evaluators should pay attention to Integrating Gender Equality Youth and Social Inclusion (GEYSI) & Gender Diversity and Inclusion (GDI) considerations across all phases of the evaluation process, irrespective of the evaluation scope and theme, recognizing that some evaluations will assess equity or inclusion as a primary focus. For guidance, consult the [CGIAR Evaluation Guidelines](#).

⁴⁵ This has been also reported by reviewers as common in much agronomy/agroecology disciplinary work.

Innovation portfolio management can be done using both hard and soft criteria. Hard criteria can include scaling readiness metrics, investment, or demand intelligence. Soft criteria can include alignment with organizational strategies, strategic niches or investment opportunities. Innovation Portfolio Management often requires information systems to support timely data collection and analysis in an evidence-based and standardized way. This type of management also requires investments in appropriate incentive, governance and resource allocation mechanisms, as well as investments in capacity and impact culture growth (Shut et al., 2024).

Annex 3. Innovation Readiness and Use Scores

Innovation readiness measures the strength of evidence that a core or complementary innovation will create positive impacts without causing undue harm in its intended context. It is a subjective score that relies on judgment and evidence, and a [calculator](#) is available to aid scoring. The readiness of enablers is scored in the same manner, as shown in **Error! Reference source not found..**

Table 6. Innovation Readiness Scores and their interpretation

Innovation readiness score	Interpretation	Description	Type of science	Type of evidence
0	Idea	Genesis of the innovation. Formulating an idea that an innovation can meet specific goal.	None	None
1	Hypothesis	Conceptual validation of the idea that an innovation can meet specific goals and development of a hypothesis about the initial idea.	Conceptual	Generic
2	Basic Model (unproven)	Researching the hypothesis that the innovation can meet specific goals using existing basic science evidence.	Conceptual	Generic
3	Basic Model (proven)	Validation of principles that the innovation can meet specific goals using existing basic science evidence.	Basic science	Generic
4	Application Model (unproven)	Researching the capacity of the innovation to meet specific goals using existing applied-science-evidence.	Basic science	Generic
5	Application Model (proven)	Validation of the capacity of the innovation to meet specific goals using existing applied science evidence.	Applied science	Generic
6	Application (unproven)	Testing of the capacity of the innovation to meet specific goals within a controlled environment that reflects the specific spatial-temporal context in which the innovation is to contribute to achieving impact.	Applied science	Generic
7	Application (proven)	Validation of the capacity of the innovation to meet specific goals within a controlled environment that reflects the specific spatial-temporal context in which the innovation is to contribute to achieving impact.	Applied science (controlled)	Specific to intervention context

Innovation readiness score	Interpretation	Description	Type of science	Type of evidence
8	Incubation	Testing the capacity of the innovation to meet specific goals or impact in natural, real, or uncontrolled conditions in the specific spatial-temporal context in which the innovation is to contribute to achieving impact with support from research efforts.	Applied science	Specific to intervention context
9	Ready	Validation of the capacity of the innovation to meet specific goals or impact in natural, real, or uncontrolled conditions in the specific spatial-temporal context in which the innovation is to contribute to achieving impact without support from research efforts.	Applied science (uncontrolled)	Specific to intervention context

Source: Adapted from Sartas et al. (2020) Table 1

Innovation use is similarly measured on a scale ranging from 0 (no use) to 9 ('common' in the livelihood system, meaning it is used under real-world circumstances by people who were not involved in its development). The scores and their interpretations are presented in **Error! Reference source not found.** These scores are also subjective and rely on judgment and evidence. The [calculator](#) supports scoring the use of innovations and enablers.

Table 7. Innovation Use Scores and their interpretation

Innovation Use Score	Interpretation	Description
0	None	Innovation is not used for achieving the objective of the intervention in the specific spatial-temporal context where the innovation is to contribute to achieving impact.
1	Intervention team	Innovation is only used by the intervention team who are developing the intervention.
2	Effective partners (rare)	Innovation has some use by effective partners who are involved in the intervention.
3	Effective partners (common)	Innovation is commonly used by effective partners who are involved in the intervention.
4	Innovation network (rare)	Innovation has some use by stakeholders who are not directly involved in the intervention but are connected to the effective partners.
5	Innovation network (common)	Innovation is commonly used by stakeholders who are not directly involved in the intervention but are connected to the effective partners.
6	Innovation system (rare)	Innovation has some use by stakeholders who work on developing similar, complementary or competing innovations but who are not directly connected to the effective partners.

Innovation Use Score	Interpretation	Description
7	Innovation system (common)	Innovation is commonly used by stakeholders who are developing similar, complementary or competing innovations but who are not directly connected to the effective partners.
8	Livelihood system (rare)	Innovation has some use by stakeholders who are not in any way involved in or linked to the development of the innovation.
9	Livelihood system (common)	Innovation is commonly used by stakeholders who are not in any way involved in or linked to the development of the innovation.

Source: Adapted from Sartas et al. (2020) Table 1

The scaling readiness of an innovation or enabler is the product of its readiness score and use score, and it ranges from 0 to 81. The overall readiness score of a package is the minimum scaling readiness score among all its components. The component with the lowest readiness score is referred to as the bottleneck because it limits the ability of the package to scale successfully.

Scaling readiness levels attach qualitative meaning to ranges of scaling readiness scores for innovations, enablers, and packages. CGIAR uses five levels of readiness: very low (scaling readiness scores from 0 to 16.2), low (16.2 to 32.4), medium (32.4 to 48.6), high (48.6 to 64.8), and very high (64.8 to 81). Readiness levels may be treated as stages of innovation development.

The scaling potential of a package is the average of the scaling readiness scores of its components. The scaling potential will always be greater than or equal to the scaling readiness score.

The scaling readiness credibility of a package is measured as the percentage of 'evidenced' innovation readiness scores and innovation use scores within a package. A score is evidenced when it is substantiated by explicitly identified sources of evidence (documented results of research, pilot studies, surveys). The total number of scores that can be evidenced is twice the number of components in a package because every component has a readiness and use score. For example, if a package is composed of six components (a core innovation, two complementary innovations, and three enablers), the total number of scores that can be evidenced is 12. If three of these are substantiated with explicitly identified sources of evidence, the evidence credibility score would be 25%.

Not all components of an Innovation Package are judged to be equally important. Stakeholders can indicate their judgments using importance ranks or importance weights. For ascending importance ranks, the most important component (the core innovation) is ranked 1. If ranks are treated as weights, then descending ranks are used to lend greater weight to more important components. Importance weights may take many forms, and ordinal weights are common. For example, stakeholders may assign weights of 1 (low), 2 (medium), and 3 (high). Weights may also reflect more granular judgments. For example, stakeholders may assign subjective weights to components that range from 0 (unnecessary) to 100 (essential).

Table 8. Scaling metrics for innovations, enablers, bundles, and packages

Score	Abbrev.	Applies to	Range	How is it calculated?
Innovation Readiness	IR	Core Innovations Complementary Innovations Enablers	0 to 9	A calculator is available.
Innovation Use	IU	Core Innovations Complementary Innovations Enablers	0 to 9	A calculator is available.
Scaling Readiness of an Innovation (or Enabler)	SRI	Core Innovations Complementary Innovations Enablers	0 to 81	$SRI = IR \times IU$.
Scaling Readiness of a Package	SRP	Bundles Packages	0 to 81	$SRP = \text{Min}(SRI)$ for all components of the package Note: $\text{Min}(X)$ means the smallest X.
Scaling Potential of a Package	SPP	Bundles Packages	0 to 81	$SPP = \text{Mean}(SRI)$ for all components of the package Note: $\text{Mean}(X)$ means the average of X.
Scaling Readiness Credibility	SRC	Bundles Packages	0% to 100%	$SRC = \frac{N(\text{Evidenced IR and IU Scores})}{2 \times N(\text{Components})}$ Note: $N(X)$ means number of X.
Component Importance Rank	CIR	Components of a Bundle or Package	Rankings	CIR = Ranked importance of a component in relation to all other components in a bundle or package. NOTE: A rank of 1 often indicates the most important component. If ranks are used as weights, then rankings may be reversed so that larger ranks provide more weight to more important components.
Component Importance Weight	CIW	Components of a Bundle or Package	Varies	Examples: CIW = 1 (Low), 2 (Medium), or 3 (High) CIW = a subjective weight that ranges from 0 (Unnecessary) to 100 (Essential).

Table 9. Breakdown of a CGIAR innovation package and metrics for the Joint Village Participatory Land Use Planning (JVLUP)

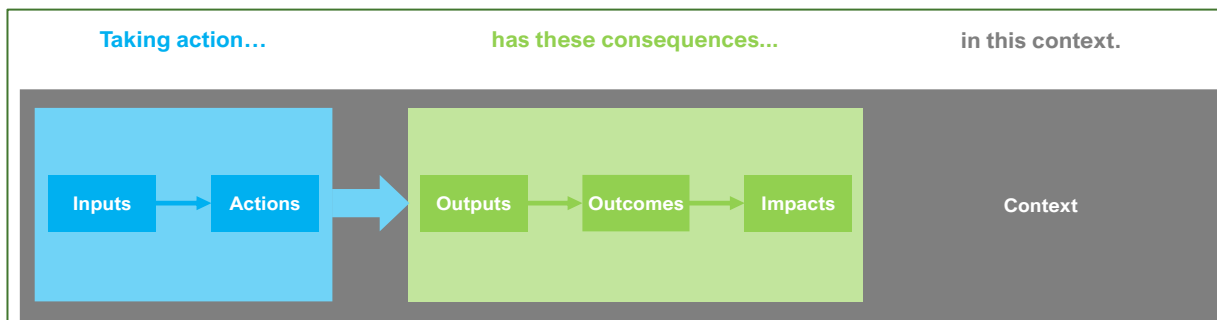
Importance	Component			Innovation Readiness (IR)	Innovation Use (IU)	Scaling Readiness of an Innovation (SRI)
	Type	Source	Description			
1	Core Innovation	CGIAR	Participatory land use planning in pastoral areas	7	4	28
2	Complementary Innovation	External	Appropriate technology for supporting planning process	1	2	2
3	Complementary Innovation	CGIAR	Platforms for inclusion of multiple stakeholders	2	1	2
4	Complementary Innovation	CGIAR	Inclusive planning process (women and youth)	5	2	10
5	Enabler (Project/ Initiative)	CGIAR	Harmonized and supporting policies and laws; regulations; and guidelines	9	7	63
6	Enabler (Project/ Initiative)	CGIAR	Adequate understanding of JVLUP and its implications	9	4	36
7	Enabler (Condition)	External	Access to accurate information to inform decision-making process during the planning phase	9	4	36
8	Enabler (Project/ Initiative)	CGIAR	Capacity building in JVLUP for both planning and implementation	9	4	36
9	Complementary Innovation	CGIAR	Living Lab that produces evidence about impacts to inform donors and investors	5	1	5
Minimum				IR 1	IU 1	2 = Scaling Readiness of a Package (SRP)
Average				Innovation Potential 6.2	Use Potential 3.2	24.2 = Scaling Potential of a Package (SPP)

Source: Adapted from Kangethe & Flintan (2022)

Annex 4. Scaling Theory of Change and Stage Transition Model

A scaling theory of change (ToC) expands the scaling thesis. It describes how scaling is expected to make impacts different than they would have been, had scaling not been undertaken or undertaken in a different way. A traditional ToC for an innovation or Program (Figure 5) is the starting point, describing how taking action (inputs and actions) will produce a specific set of consequences (a cascade of outputs, outcomes, and impacts). This process is situated in a context, often viewed as a system, that may affect every part of the process.

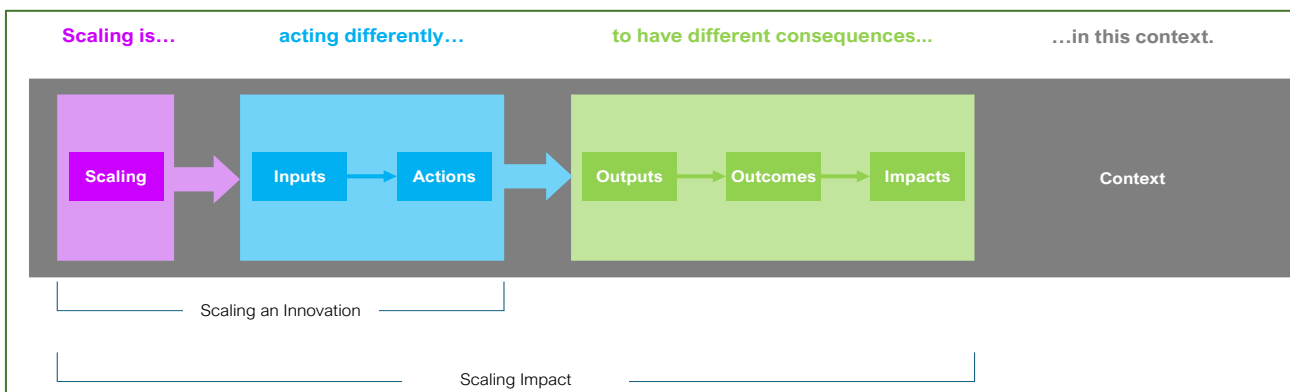
Figure 5. ToC of an innovation



Source: John Gargani

A linear (static) scaling ToC (Figure 6) describes how scaling changes the way scaling actors act, and how this in turn creates different consequences. **Scaling an innovation** refers to how scaling actors behave differently when they scale in different contexts. **Scaling impact** refers to the entire contextualized process. This type of linear logic is often associated with a 'bigger is better' view of scaling and is perhaps the most common mechanism assumed in scaling ToCs. While linear logic may be sufficient in some cases, it presumes a static relationship between scaling and context, such that neither meaningfully influences the other as scaling unfolds.

Figure 6. Scaling ToC, scaling an innovation and scaling impact

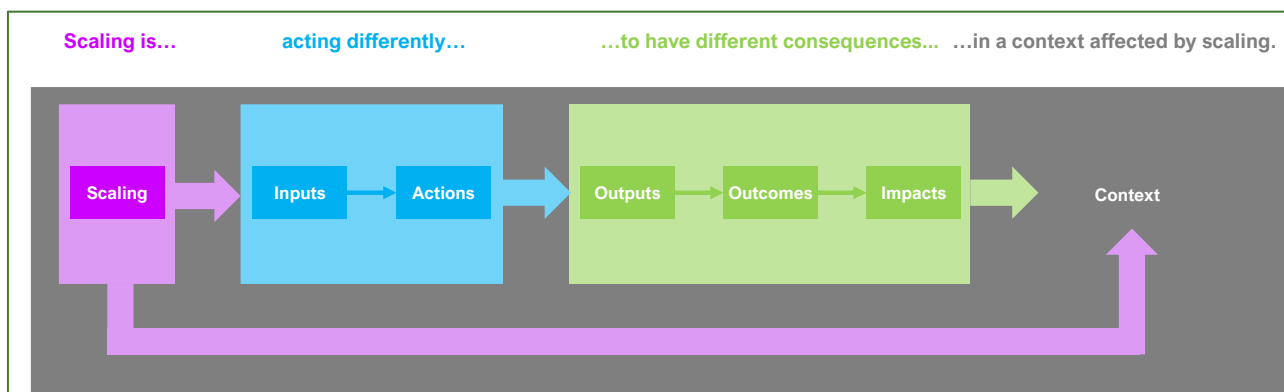


Source: John Gargani

A non-linear and dynamic scaling ToC considers how scaling and the context system interact. First, if scaling is successful, outputs, outcomes, and impacts may change the context (Figure 7). For example, an innovative fertilizer that increases cassava yields may increase supply so much that prices fall, changing the market context. Second, scaling itself may change the context. For example, funders may consider the new fertilizer a good social investment to the point that it crowds out funding for other worthy innovations in the region. Third, as context changes, it has the potential to change every part of the causal process, making it more or less likely that positive impacts will be realized. For example, falling cassava prices would make it less likely that farmers will plant cassava and more likely they will switch back to their previous crops that do not benefit from the fertilizer. This would in turn make it less likely that the environmental benefits of the new fertilizer will be realized. This interaction of scaling and context is what makes this scaling ToC dynamic and non-linear.

Evaluators can help scaling actors develop, update, and test their scaling ToC, but it should not be structured exactly like a ToC for an innovation or program, and instead should build on it.⁴⁶ In CGIAR, scaling actors may choose either a linear or non-linear scaling ToC, and they should be able to justify their choice.

Figure 7. Considering context in a scaling ToC



Source: John Gargani

Stage transition models

Stage transition models document how innovations move back and forth between stages of development. Figure 1 in [CGIAR’s 2023 Annual Report](#) includes a graphic illustrating how 376 innovations moved between stages over a twelve-month period. The same information is presented in a simpler transition matrix (Figure 8) that may be used to calculate scaling metrics.

⁴⁶ The [MELIA Summary](#) can help provide further evidence on ToCs in CGIAR.

Figure 8. The underlying stage transition matrix

		To End of Year					Start of Year	
		Very Low	Low	Medium	High	Very High	Total	Percent
From Start of Year	Very Low	11	11	4	4	4	34	9%
	Low	0	42	25	20	1	88	23%
	Medium	0	0	63	14	6	83	22%
	High	0	0	7	55	19	81	22%
	Very High	0	1	2	7	80	90	24%
End of Year								
Total		11	54	101	100	110	376	100%
Percent		3%	14%	27%	27%	29%	100%	

Source: John Gargani

Some of the innovations increased their readiness during the year and advanced to a higher stage (the green cells), while some decreased their readiness and moved to lower stages (the blue cells). The majority remained at the same stage of development (the grey cells). Overall, 29% of innovations advanced to a higher level of readiness by year end, 67% remained unchanged, and 5% decreased their level of scaling readiness. These are descriptive measures that, in the absence of standards, do not support evaluative conclusions.



Independent
Advisory and
Evaluation
Service

Independent Advisory and Evaluation Service

Alliance of Bioversity International and CIAT

Via di San Domenico, 1 00153 Rome, Italy

IAES@cgiar.org

<https://iaes.cgiar.org/>