MEL ASSESSMENT SOIL VALUES

REVIEW

seo • amsterdam economics

AUTHORS
ROBBERT RADEMAKERS, DEBBIE KEIJSER-SMIT & NIENKE OOMES

COMMISSIONED BY
DUTCH MINISTRY OF FOREIGN AFFAIRS (MFA)

AMSTERDAM, FEBRUARY 26

SEO note 2024

Information & Disclaimer
SEO Amsterdam Economics has not performed any research on the obtained information and data that would constitute an audit or due diligence. SEO is not responsible for errors or omissions in the obtained information and data.

Roetersstraat 29
1018 WB, Amsterdam
The Netherlands

+31 20 399 1255
secretariaat@seo.nl
www.seo.nl/en/

Introduction

Soil Values is a EUR 100 million, 10-year programme that aims to improve soil fertility and farm productivity. The programme started in 2023 and works through a portfolio of partnerships that co-design and implement context-specific solutions for sustainable soil and fertiliser management. The programme operates in selected countries in Sub-Saharan Africa and seeks to reach large numbers of farmers, both directly and indirectly. It focuses on strengthening input systems, agronomic practices, and enabling services.

The programme is expected to track results at both partnership and programme level. Each partnership is asked to develop a partnership-specific Theory of Change with causal pathways and indicators. Data will be collected through partner reporting, surveys, and qualitative methods. Results are aggregated at programme level to track progress.

SEO Amsterdam Economics was commissioned by the Dutch Ministry of Foreign Affairs (MFA) to assess the quality and practical applicability of the Soil Values Monitoring, Evaluation and Learning (MEL) system. The primary purpose was to determine whether the MEL system is fit for purpose—i.e., whether it can credibly track key envisaged results using clearly defined and measurable indicators, and whether it can support valid contribution and, where relevant, attribution claims. The assessment also considered whether the system is likely to support learning and inform programme steering and evaluation. A secondary purpose was to develop and pilot a framework that can be used to assess other MEL systems in the future.

Methodology

To assess the Soil Values MEL system, SEO developed a structured scoring framework in collaboration with MFA. The framework is based on eight pre-defined criteria that reflect international best practice in programme-level MEL design and implementation:

1. Coherence of the MEL system with the Theory of Change (ToC);
2. Relevance and quality of the indicator set;
3. Quality of measurement methodologies and the underlying evidence base;
4. The learning and steering function of the MEL system;
5. Accountability towards BZ/DGIS, including alignment with FNS results frameworks;
6. Linkage to mid-term and end-term evaluations in line with OECD-DAC criteria;
7. Practical feasibility of the MEL system.

For each criterion, SEO formulated a set of indicators to assess to what extent the MEL system meets minimum requirements and supports effective monitoring, learning, accountability, and evaluation.

The assessment was based on a desk review of key programme documents provided by the MFA. These included the MEL Strategic Guidance Note; an Excel file with an overview of the indicator definitions and calculation methodologies; the Partnership Guide in Excel; the Geoportal note for the Annual Work Plan and Budget; the document on measurement and aggregation of Food and Nutrition Security (FNS) results (2025); and the full proposal for the Soil Fertility Grant Programme. Findings were triangulated across documents to assess internal consistency and to identify strengths, gaps, and areas requiring further clarification.

Assessment

1. Coherence with the ToC

Quality of ToC (1.1): The programme's current ToC provides a detailed overview of the programme's inputs and activities and envisaged results (outputs, outcomes, impact). It has a high information density (targets, actors, pathways, intervention areas) and reads more like a condensed programme description than a ToC that can be used for MEL purposes. Based on common quality criteria, this assessment identifies three areas for improvement.[1]

First, the **causal links** in the ToC ("arrows") are not sufficiently explicit: many arrows connect most boxes, making it difficult to derive testable hypotheses about the programmes presumed contribution and attribution. In addition, ToC levels imply a time sequence, so causal arrows should run only upward–from lower-level results to higher-level results (e.g., outputs to outcomes).

Second, given the indirect approach of the programme to achieve results through partnerships, the ToC should more clearly distinguish between activities (what the programme does directly); outputs (results that are in the sphere of control); and outcomes (results that are expected to be achieved indirectly, through partners). Several indicators that are currently listed as outputs in the ToC could be considered outcomes that are part of the sphere of influence but not control. The ToC also does not clearly specify activities (labelled as "intervention areas") and does not summarise inputs (financial and human resources).

Third, key assumptions are not stated explicitly; some are implied (e.g., availability of fertilisers) but not formulated as trackable conditions, nor linked to mitigation measures. However, the MEL strategy does include a detailed plan to link context change to adjustments to the ToC, with specified mechanisms and frequency (half-yearly context/conflict analyses feeding into a biannual ToC review). The MEL framework foresees more detailed partnership ToCs[2] that could address causal logic and assumptions, but provides limited guidance nor clear examples which risks variation across partnerships and reduces comparability and aggregation at programme level.

Alignment of the MEL system with the ToC (1.2): The set of indicators (Annex 1 of the Strategic MEL guide) makes a clear distinction between output and outcome variables. However, the listed indicators are not explicitly linked to variables, causal links (the arrows in the ToC), or the pathways within the ToC. Finally, while the programme is implemented through partnerships, the indicator set on does not yet reflect this structure consistently. In particular, output indicators should include only what the programme delivers directly to partnerships (e.g., outreach and capacity strengthening), i.e, what is directly within the sphere of control of the programme.

2. Relevance and quality of indicators

Relevance and coverage (2.1): Relevance and coverage (2.1): The MEL framework includes a detailed and broad set of food and nutrition security (FNS) and programme-specific indicators, covering several parts of the intervention logic and multiple pathways. However, relevance is weaker in some areas. Particularly, current indicators do not cover what the program directly delivers to partnerships ("outputs", see 1.2). In addition, no impact-level indicators are defined, and the system does not yet include a structured approach to tracking externalities, particularly negative ones, and monitoring whether key assumptions hold. Finally, indicators are not clearly organised by the three

1 Theory of Change (BetterEvaluations and UNICEF, 2014)
2 MEL Strategic Guidance Soil Values p.9

pathways (agro-ecological, inclusive market, enabling environment), and it is not specified which indicators are disaggregated and for which groups[3]; an important gap for MFA reporting.

Measurement (2.2): The MEL framework includes a detailed measurement plan, and Annex A clarifies what indicators measure and summarizes calculation methods. However, not all variables listed in Table 1 are explained in Annex A[4] and some indicators are not yet operationalized in a way that ensures consistent measurement.

First, not all indicators are operationalized SMARTLY. Important examples are income and productivity. The current measurement of income does not specify how it will asked (in surveys) and calculated (e.g., reference period, seasonal adjustment, treatment of in-kind income, or how outliers and recall bias will be handled); neither does it mention any mitigation strategies for common biases[5]. For some indicators, key elements are not well defined, which makes indicators hard to operationalise—for example, what counts as an "older person," an "ethnic minority," or a "cross-sector partnership". For the FNS-related indicators, the definitions and calculation logic align with MFA methodological notes. Furthermore, the MEL framework has not set any targets;

Additionally, methodological guidance is limited on other elements. For instance, there is no explanation on how the tiered counting approach (EOP) will be applied and how double counting is prevented. Without clear partnership and geographic boundary definitions and a standardised counting protocol, the system is vulnerable to double counting and unsupported extrapolation of results across locations and partners. The MEL note indicates that programme teams will define partnerships and geographic boundaries ex post (once the partnerships are established), but greater clarity ex ante would support program design and improve consistency from the outset.

The use of geospatial data is mentioned but not yet part of the list of indicators (Annex A), so this component cannot be assessed at this stage.

3. Measurement methodology

The measurement methodology has several relevant building blocks in place: a baseline is conducted, the MEL framework distinguishes where attribution may be feasible versus where contribution approaches are more appropriate, and the data management system is described including definitions and calculation methodologies (see Annex A). However, the current methodology is not sufficiently well developed to support credible attribution claims (indicator 3.2).

Data collection and data management (3.1): The programme sets out a clear data management system (Section 6.6) and proposes a mixed-methods approach, which provides flexibility. At the same time, the data management plan could be strengthened by specifying how datasets and documentation will be stored, anonymised, and shared to facilitate future external evaluations (see section 6).

Soil Values has already conducted a sample baseline (3,580 farmers) in nine watersheds targeted by the programme. However, the baseline appears to cover only targeted watersheds, which makes it difficult to establish

3 Some forms of disaggregation are mentioned in Annex 2 (Partnership guide), but the document does not operationalize what counts as "youth" nor other disaggregation variables. For clarity and consistency, the MEL team could align this with the DGIS definition: youth are individuals under 35 years of age (i.e., 0–34)

4 e.g., Percentage increase in food production (kg) or yields at household level)

5 For example, if you ask households about last year's income, they may simply extrapolate from their current situation; if you ask about last month's income and annualise it (multiply by 12), you risk distortion due to seasonality.

a valid control group and prove attribution (see 3.2). Furthermore, not all indicators are measured at baseline, which makes it difficult to demonstrate change over time and limits the scope for attribution (see Annex A).

Furthermore, the sampling approach is not yet sufficiently substantiated (for the baseline and other primary data collection) and relies on "naïve" assumptions. There is a risk that the current sample (including the baseline survey sample) underestimates the number of observations needed to detect meaningful change and support attribution or contribution.

A key issue is that the primary unit of variation is not identified and used to design the sampling strategy. In programmes like this, the main unit of variation is often the village (and not the household): within a watershed or geographic area, some villages receive an intervention (or a higher intensity of intervention) and others do not. In that case, the village becomes the main unit of analysis and sampling plan needs to be designed accordingly, since households within a village are typically all exposed (or not exposed), leaving little meaningful within-village variation. Nearby villages may also influence each other (contamination/spillovers), so credible attribution requires a sufficient number of both treated and non-treated villages across treated and non-treated watersheds.

The sampling design also requires specification on other issues. Key assumptions for power calculations to estimate the required sample size are not fully justified (e.g., expected change, outcome variance, non-response/attrition, panel vs. repeated cross-section), the sampling frame and eligibility criteria per partnership/area are not described in detail, and there is no clear plan for stratification or oversampling (notably for women, minorities, and youth) and the corresponding use of weights in analysis. In addition, sample size calculations should account for likely clustering within partnership geographies (design effects/intra-cluster correlation), which may require a larger sample to maintain statistical power.

Finally, although the plan references "ethical data assets," it does not yet spell out privacy safeguards and "do no harm" procedures—particularly important given that data collection and implementation run through partnerships.

Contribution and attribution strategy (3.2): In its current form, the MEL framework is unlikely to establish attribution or to demonstrate contributions in a clear and testable way. With minor and relatively straightforward adjustments (e.g., explicit causal hypotheses, clearer use of evidence across data sources, and a basic synthesis approach), estimating contributions should be feasible, for example in the form of 'contribution scores', using a form of Contribution Analysis (see Annex A for example). Establishing attribution, however, would require more structural design changes that are more complex and costly. At present, the framework relies mostly on qualitative evidence and mixed methods (interviews/FGDs), including reference to the "BACO approach". While qualitative methods are valuable for understanding mechanisms and triangulating findings, they cannot rigorously establish attribution because they do not allow estimation of a treatment effect (relative to a counterfactual),[6] and the BACO approach is not sufficiently explained—particularly how the counterfactual will be defined and assessed.

The planned quasi-experimental impact survey (difference-in-differences, DiD) can potentially support attribution claims, but only if key identifying assumptions are met; in particular, the presence of parallel trends and the absence of spillovers/interference (SUTVA). Based on the current design, several assumptions are difficult to defend, which increases the risk of biased estimates:

[6] Ter Weel, B., Janssen, M., Bijlsma, M., & De Boer, P. J. (2022, oktober). *Durf te leren, ga door met meten: Op zoek naar kaders en methoden voor de evaluatie van systeem- en transitiebeleid* (Publicatienummer 2022-115). Ministerie van Economische Zaken en Klimaat.

- **Parallel trends:** Treated and comparison areas (or households) should be sufficiently similar to justify the assumption that, absent the programme, they would have followed similar trends. Because the baseline appears to cover only targeted watersheds, it is difficult to identify truly comparable untreated areas. This raises risks of selection-on-trends and time-varying confounding (e.g., different shocks, other projects, market changes), which can bias estimates in either direction. Furthermore, the MEL system does not include a methodology that would support the validity of this assumption (e.g., advanced propensity score matching, ex ante balance checks, etc.).

- **No spillovers/interference:** Outcomes in control areas should not be affected by treatment. With geographically close watersheds and partnership activities that may spread (knowledge, practices, inputs), contamination is plausible. This can dilute differences between treated and control areas and undermine the counterfactual.

- **No anticipation / no pre-treatment effects:** If farmers or partners change behaviour in expectation of support before "treatment" starts, baseline-to-follow-up changes will be distorted, biasing results.

- **Sufficient independent treated units:** With only two treated partnerships, statistical leverage is very limited. Effects become hard to detect, uncertainty is unstable, and results may be driven by idiosyncratic changes in a single treated unit.

- **Correct unit of variation (see 3.1):** Treatment and control should be defined at the level where exposure actually varies (often villages/communities). If exposure is largely uniform within villages, the effective sample size is the number of villages, not households; ignoring clustering can overstate precision and understate the required sample size.

While assessing contributions is more feasible than assessing attribution, the current MEL framework does not yet specify a clear methodology (e.g., Contribution Analysis) to combine and assess evidence across data sources, including how different sources will be integrated to support contribution claims for key outcomes. To make contribution claims credible, this needs to be operationalised: for each key outcome, the framework should set out the main causal hypotheses, the types of evidence expected (quantitative and qualitative), and clear rules for how evidence will be triangulated and weighed (including how alternative explanations will be tested and addressed). Furthermore, the contribution methodology should explicitly address the reliability and strength of the evidence (for example by applying the Effectladder developed by the Ter Weel Committee, as illustrated in Annex A).

Validation, aggregation and quality assurance (3.3): The Strategic MEL note includes several validation elements through a mixed-methods approach, aiming to triangulate evidence from interviews, annual validation surveys, and partner reporting. The planned MIS and partnership documentation (including means of verification and calculation sheets) also support traceability of reported results. However, quality assurance is still described in general terms: the note mentions training, supervision and "quality checks," but does not yet set out clear procedures (e.g., back-checks, call-backs, plausibility/outlier rules, GPS/time-stamp checks, data-cleaning logs, and escalation steps).

The guidance note describes a bottom-up aggregation logic (partnership → watershed → country → programme) and appropriately recognises the need to avoid double counting when results overlap. It also proposes a tiered approach to estimate outreach/results (MIS counts where possible, complemented by partner estimates and extrapolation from validation surveys). However, the aggregation and extrapolation rules are not yet sufficiently

detailed for consistent use across partnerships and countries, and there is no standard calculation template to make roll-ups comparable and auditable.

4. Learning and steering

Overall, the Soil Values MEL system provides some foundations for learning and steering, but key elements are not yet sufficiently specified. While the MEL strategic guidance note sets out programme-wide learning questions, it is unclear how these have changed over time and, crucially, how data and analysis will be used to answer them. The documents do not describe fixed learning mechanisms or learning loops with clear roles for who collects, analyses and decides, nor do they explain how lessons will be documented and shared, beyond a general ambition to mainstream soil fertility through collaboration with regional and international partners.

Steering and governance are more developed, but still partly unclear. Governance arrangements refer to an in-house MEL team and the use of local youth enumerators, yet responsibilities and decision lines remain ambiguous across partnership and programme levels. Adaptive management is strongly articulated in the proposal, particularly in relation to context and security risks and the ability to reallocate activities and budgets, but this is not clearly embedded in the MEL guidance note. While capacity building plans are specified for youth enumerators, the extent of MEL capacity building for implementation teams and partners is not clear.

5. Accountability towards MFA

The Soil Values documentation clearly specifies what will be reported to MFA/DGIS and how often. For all aggregation indicators, it is stated whether an indicator is a DGIS indicator, alongside the methodology, measurement and reporting frequency, and whether results are measured as contribution or attribution. The set includes three DGIS output indicators and eight DGIS outcome indicators, and the indicator definitions and methodologies are consistent with MFA guidance. However, the documentation does not clearly explain how the data will ultimately be packaged and presented to DGIS (e.g. in which specific reporting products and formats).

6. Evaluation readiness ('evaluability')

The MEL documentation provides some foundations for external evaluation, but it is not yet fully structured around OECD-DAC requirements. The guidance note mentions that monitoring will generate a database to inform future evaluations, and evaluability is supported through partnership Measurement Plans and systematically storing data in a Monitoring Information System (MIS). The system also explicitly aims to generate insights into effectiveness and includes measures of efficiency (e.g. costs per beneficiary reached or area covered). However, OECD-DAC criteria are only referenced in general terms and the documents do not present a complete OECD-DAC evaluation framework across criteria, with key elements (including indicators) still subject to review after ILMPs. In addition, while an Independent Mid-Term Review is planned, there is no documented plan for ensuring independence and quality assurance of external evaluations (e.g. role division, review/validation steps, conflict-of-interest management, or transparency about limitations).

7. Feasibility

The MEL system appears broadly proportionate to the scale and complexity of Soil Values. Using a benchmark of 3-7 percent of programme budgets that are typically allocated for MEL implies a MEL budget of around €3–7 million, likely towards the upper end given indirect pathways and complexity. The system also reflects the programme's risk context through elements such as biannual conflict/context reviews and biannual ToC updates. However, the actual MEL budget is not clearly stated, and proportionality is difficult to confirm because the number of partnerships (a

key workload driver) is not yet known and because, as the guidance note cautions, the MEL system could become too complex and heavy.

Capacity and feasibility are partly addressed in the MEL design. Roles are defined across levels, and it is valuable to have a dedicated MEL team member per partnership and a central MEL team responsible for annual survey design and triangulation of evidence. We also value the plans to recruit youth from partnership areas as enumerators and provide training, close supervision and quality checks. Capacity gaps are identified for youth, but not clearly for all implementing teams. Practical constraints are considered implicitly through the use of local enumerators and some attention to seasons and risks, but access, safety and logistics are not yet explicitly translated into MEL operational planning or timelines.

Recommendations

The documentation would benefit from a single, consolidated overview of the MEL system. At present, key design and operational details are split across the full proposal and the MEL Strategic Guidance Note, and each document adds new information rather than providing one coherent end-to-end description. This makes it harder to understand the complete MEL architecture, trace roles and workflows, and assess feasibility and evaluability. SEO therefore recommends developing one integrated MEL overview document that brings together the ToC structure, indicator framework, data sources and responsibilities, reporting products and timelines, learning and governance mechanisms, and how the system will be adapted over time.

Avoid postponing key MEL design decisions. The programme currently leans heavily on partnership-level inputs and intends to finalise several core MEL elements once partnerships are established. However, if foundational choices are deferred too long, it becomes difficult to develop a robust (quasi-)experimental design, ensure consistency in MEL practices across partnerships, and prevent later misalignment between indicators, and data collection tools. Table 1 therefore sets out recommendations to support timely decision-making and help lock in core MEL elements early.

Specific recommendations per criterion are presented in the table on the next page.

Table 1 Main recommendations per criterion

Criterion	Main recommendations
Coherence with TOC	• Simplify and sharpen the programme ToC with explicit upward causal links (see Annex A for example). • Clarify spheres of control vs influence by separating what the programme does directly (inputs/activities/outputs to partnerships) from results expected through partners (early outcomes). Include key assumptions and externalities as additional elements of the ToC. • Standardise partnership ToCs with a clear example and better defined building blocks (inputs → activities → outputs → outcomes, assumptions/risks + mitigation, pathways), aligned to the programme-level ToC to improve comparability and aggregation. • Link indicators directly to the ToC and causal pathways by adding tags for each indicator: *ToC box/variable*, *arrow (causal link tested)*, and *pathway* (agro-ecological / inclusive market / enabling environment).
Relevance and quality of indicators	• Strengthen "output" indicators to reflect what the programme delivers to partnerships (sphere of control), not what partners achieve. Add a coherent output bundle covering: outreach, convening, technical assistance, capacity strengthening, MEL support, and adaptive management. • Introduce a light "impact-level" layer (a few headline indicators) aligned with programme ambition and MFA reporting, with clear positioning as contribution-oriented where attribution isn't feasible. • Make explicit which indicators will be disaggregated and at which level (women, youth, etc.). Adapt the sampling plan accordingly (see measurement methodology). • Operationalize indicators more consistently—a practical way is to start drafting survey and interview guides early and tighten definitions of key concepts (e.g., ethnic minority, youth, income, productivity). Integrate geospatial data: it has high potential, but clarify what it will be used for (targeting, verification, exposure/intensity, land use/soil proxies) and reflect this in the indicator set.
Measurement methodology	• Prioritize a clear contribution methodology (Contribution Analysis) to synthesise data sources: define the main causal hypotheses per outcome, specify expected evidence (quant + qual), and set explicit rules for triangulation and testing alternative explanations (see Annex B for an example). Where possible, define standard "contribution scores" (with clear criteria) and report the degree of certainty alongside each claim using the Effectenladder (Commissie ter Weel, see note 5). • Strengthen the attribution strategy where feasible by expanding the pool of treated and non-treated villages for comparison, and applying robust matching/weighting approaches (e.g., propensity score / covariate matching) combined with DiD where assumptions can be defended and contamination managed. • Rework sampling around the correct unit of variation (often the village) and update power/sample-size calculations accordingly (ICC/design effects, expected change/variance, attrition, panel vs repeated cross-section). Include a clear plan for stratification/oversampling (women, youth, minorities) and weights. • Fix baseline and panel feasibility gaps: ensure priority outcome indicators are measured at baseline (or introduce a phased baseline/top-up where needed), and document clear treatment timing to avoid anticipation/pre-treatment bias.
Learning and steering	• Establish a formal learning and steering cycle in which the MEL team produces a concise data-driven learning brief linked to evolving learning questions, followed by a structured reflection meeting with clearly assigned roles and a documented decision log to ensure that findings systematically inform programme adjustments. • Develop a (MEL) capacity building plan implementation teams and partners.
Accountability	• Elaborate in the MEL guidance note how data will ultimately be packaged and presented to DGIS.
Evaluation readiness	• Ensure that the MEL system systematically collects and stores data aligned with all relevant OECD-DAC criteria by clearly linking existing indicators, monitoring tools and MIS data fields to relevance, coherence, effectiveness, efficiency, impact and sustainability, so that external evaluators can readily access complete, structured and traceable evidence across all criteria. • Develop a plan for ensuring independence and quality assurance of external evaluations (e.g. role division, review/validation steps, conflict-of-interest management, or transparency about limitations).
Feasibility	• Given the complex nature of this programme, assign at least 5 percent of the total programme budget to MEL activities. • Conduct a structured MEL feasibility assessment and translate identified gaps and operational constraints—such as access, safety, seasonality and logistics—into a realistic MEL operational plan with adjusted timelines, roles and mitigation measures to ensure feasible and consistent data collection.

Appendix A Example contribution analysis table

Table A.1 Example contribution methodology

Observed change	Factor	Type	Evidence	Contribution score (see Table A2)	Strength of the evidence (see Table A3)
Improved income (= indicator from ToC)	Adoption of balanced fertiliser + lime (via grant)	Primary	Interviews, surveys, FGDs		
	Improved farmer knowledge on soil testing & application rates	Primary	Interviews, surveys, FGDs		
	Good rainfall season	Contributing data	Survey, geospatial data		
	Increase in fertiliser market prices	Contradicting rival factors	Desk research, interviews		

Table A.2 Example of contribution scores

Level	Strength of contribution	Description
1	No contribution	No credible evidence Soil Values influenced the change; results reflect other factors.
2	Weak contribution	Limited influence in a small part of the change; evidence is thin and/or mostly indirect. Other contributing or contradicting factors are major driver.
3	Moderate contribution	Clear influence on several key steps/results; other factors also important; evidence is reasonably credible.
4	Strong contribution	Soil values is a major driver of the change; strong, consistent evidence across sources; plausible alternatives largely ruled out

Table A.3 Ter Weel's Effectenladder to indicate strength of evidence

Effect score	Name	Type of research	Wording used in conclusion
5	Causal	Experiments (gold standard); natural experiments combined with econometrics	Proven (in)effective
4	Plausibly causal	Combination of descriptive, promising and indicative	Very likely (in)effective
3	Indicative	Meta-analyses[2]; other econometric analyses[3]; demonstrate goal achievement, trends or behavioural change	Likely (in)effective
2	Promising	Expert judgement; policy theory; theoretical models; theoretical analyses; statements about effectiveness in surveys and interviews	Expected (in)effective
1	Descriptive	Description of the goal, the target group, the prerequisites and the intervention method	Possibly (in)effective

Appendix B Example Theory of Change