Assessing standardized SDG contributions at the project level
Practical application and concepts

Robert Müller* & Thijs Merton
*Contact: robert@bridge-builders.de

About this document

This document is based on an independent research that intends to create a theoretical foundation for measuring project results across disciplines. The guidance provided here is meant for practical application. The document therefore starts by providing an overview, while detailed concepts are described in section 4. Readers who are mainly interested in the practical application of SDG-indicators may also have a look at Annex 2, before reading through the entire document.

The approach described here draws on existing concepts of impact evaluation, theory of change, result-based finance, impact investment, certified GHG reduction and sustainability assessments of carbon projects. It represents an attempt to integrate the experiences gained in these different fields. One of the main guiding questions is if and how the standardization that is common in the field of climate change mitigation, where project outcomes are standardized and even traded as tons of CO_2-equivalents, can be applied to other areas.

The concepts presented here are work in progress. We are grateful for all kinds of feedback and input.

The lead author received funding by the Andrea von Braun Stiftung for interdisciplinary research (www.avbstiftung.de). The research was conducted in close cooperation with Bridge Builders (bridge-builders.de; Ole Meier-Hahn, Claudia Schonter and Sven Bratschke) and the FairClimateFund (www.fairclimatefund.nl; Neera van der Geest and Gert Crielaard).

Furthermore, it was supported by several institutions and individuals, including Aqua for All, Xaver Kitzinger (Kaffeekooperative.de), Christine Haida (Aussichtsreich e.V.), Dr. Stefan Kreft (Centre for Econics and Ecosystem Management, Eberswalde University for Sustainable Development), Prof. Dr. Bernd Hansjürgens (Helmholz Centre for Environmental Research – UFZ) and Prof. Dr. Inga Schleip (Eberswalde University for Sustainable Development).

Table of contents

1 Introduction and summary 2
2 Systematic assessment: Can project-level SDG contributions be standardized? 3
3 Application of the assessment criteria to different categories of SDG targets 4
4 Explanation of concepts 6
***5* Conclusions** 11
Annex 1 Examples of selected project-level SDG indicators in practice 12
Annex 2: List of potential project-level indicators per SDG 16

1 Introduction and summary

Progress towards SDG targets is measured globally through official SDG indicators. These are suitable for regular (annual) observations at the country-level. But official SDG indicators are often difficult to apply to projects and programmes, due to two reasons:

- Official SDG indicators mostly measure relative change, for example, in reference to a country's population. But a project's SDG contributions depend on its size, for example, the size of its target group. Comparisons between projects require standardized units that represent absolute values. These units also need to consider the duration of a project's effects.
- Official SDG indicators observe the cumulative effects of different kinds of policies, activities and dynamics, without a primary need to know about causalities. But measuring a project's results requires explaining how this project is responsible for the observed change, among others, by comparing with a baseline or counterfactual.

In order to cope with these challenges, a systematic approach for project-level SDG measurement is defined, drawing on impact evaluation, theory of change, experiences from result-based finance, impact investment, certified GHG reduction and sustainability assessments of carbon projects.

The approach strives for a profound understanding of project results to be measured, which includes assessing the intrinsic values behind the indicators and the level where change occurs (output-outcome-impact, see below). The approach is not to be seen as a straight instruction on how to derive project-level SDG indicators, but rather as an impulse for reflection about measuring project results.

It is found that human wellbeing can be standardized quite well by referring to the universally equivalent value of human life or its quality. Some of these may be quantified in "person-years", meaning that the value of an achievement (like access to drinking water) increases linearly with the time it persists. This implies a need for regular systematic ex-post monitoring.

For SDGs targeting nature, achievements are more difficult to standardize, among others because nature is mostly targeted indirectly by mitigating threats, making it difficult to trace results back to a project.

We hope that applying the suggested procedure can increase clarity on change assessment and support result-oriented project planning, even in the many cases where change is difficult or impossible to quantify.

Example of a standardized, project-level SDG indicator measured in person-years:

Two projects promote the dissemination of solar home systems (SHS) in areas without grid connection. Both distribute of 200 units, each serving 5 persons in average. After 4 years, it is found:

Project A: In average, the usage rate of SHS was 50%.

Project B: the usage rate of SHS was 100%. But after three years, the national grid was extended to the project region. Therefore, households stopped using the SHS.

How could the projects' contributions to SDG 7 be measured (original indicator: Proportion of population with access to electricity? Should project B be regarded worthless, since it stopped? Or should it be rated higher due to the better adoption of the technology?

We suggest to use the unit "person-years", considering the permanence of results:

Project A achieved 2,000 person-years of electrification (50% of 1,000 persons, for four years).

Project B: 3,000 person years of electrification (100% of 1,000 persons, for three years).

Monitoring after 8 years might show a higher value for project A, since it is still ongoing. While perfect standardization will never be reached in a complex world, this approach allows at least to consider the dimension of time, which is crucial if evaluating the results of projects.

2 Systematic assessment: Can project-level SDG contributions be standardized?

To be used to define new standardized indicators, or as an inspiration to improve existing indicators

→ ***Detailed explanations of concepts are found in section 4!***

Step 1: Analysis of the value to be measured

1.a) Identify the **underlying intrinsic value,** whether it has universal **equivalence,** and the **category** of the approach taken to pursue SDGs (see section 3 for details on categories).

Human life and **life-years with a specific quality** (universally equivalent, intrinsic values).
Categories: Basic services; Personal resources; Risk Protection; Rights & Participation

Components of nature and environment. Intrinsic values are assumed for species, ecosystems or ecosystem services, but no universal equivalence as for human life.
Categories: Specific threat prevention; protection; direct ecosystem management/manipulation

1.b) Determine the **levels of change** to be measured:

Indicators can measure *inputs, outputs, outcomes* or *impacts* of a project. This *impact pathway (theory of change)* can also be seen as a continuous spectrum from *means level* to *ends level.* In a broad sense, *means* can also refer to the mitigation of threats to ecosystems, for example.

Awareness about levels of change can be very helpful for measuring change effectively.

Step 2: Definition of a standardized unit *(if possible)*

2.a) Assess the **scale of measurement**: Only **absolute values** make projects comparable ("total persons with …" instead of "share of population with…"). This works best with continuous scales (US$, hours), or dichotomous scales (yes/no) summed up "number of persons with…". **Dichotomy** is often a simplification and requires **threshold definition** (poverty level, basic service level, …)

2.b) Assess **Permanence**: Some achievements are automatically durable (such as literacy), others can be lost again (such as access to education). This needs to be considered for quantifying results; and it implies a need for regular ex-post quantification of non-durable achievements.

2.c) Assign a standardized unit

Three types of units can be used to quantify a large range of project-level SDG indicators:

- ***Person-years*** *with a defined minimum achievement* measure non-permanent, dichotomous results (such as person-years with basic drinking water access). Person-years are defined as equivalents: Two persons with access for one year = one person with access for two years.
- ***Number of persons*** *with a defined minimum achievement* apply to permanent, dichotomous results, such as people vaccinated or people able to read
- Total ***US$*** earned, ***hours*** gained, ***ha*** reforested or *conserved*, ***ton****s of GHG avoided, etc.* apply to continuous results and can totalize a project's achievements.

2.d) Assess general quantifiability: Quantifiability varies for different types of SDG targets and for different levels of change. Targets that are less quantifiable are therefore not less important.

Step 3: Measurement of change

3.a) Establish the baseline or **counterfactual, a** non-project scenario against which change is measured. An ideal baseline is **independent** of the project and **dynamic**, considering changes after the project start. Both requirements are given if comparing with an independent control group, which is however often hard to realize in practice).

3.b) Assess attribution, defining the share of change for which a project is responsible. This is usually much easier at the *means- level* than at the *ends-level*, due to growing influence of external factors.

3.c) Define the **Means of Verification.** Different SDG targets can be verified in different ways, such as: Direct observation of physical change, surveys/interviews, standardized tests, proxy-indicators or modelling. The reliability of measurements depends on the means chosen for verification.

Step 4: Quality control

4.a) Identify potential biases (selection bias, courtesy bias, framing, confusion, deficient indicators, ...)
4.b) Identify potential risks and unintended effects: Systematic screening.
4.c) Assess the reliability of quantified results: Distinguish "verified", "estimated", "narrative" results)
4.d) Equally consider non-quantifiable achievements

3 Application of the assessment criteria to different categories of SDG targets

3.1 Human Wellbeing: The categories defined below represent different approaches of pursuing **human wellbeing**, with common characteristics of measurability. The table may be used to orient the assessment of standardized indicators. → ***Detailed explanations of concepts are found in section 4!***

1.a) Category of SDG target referring to: Human Wellbeing	**Basic services and opportunities:** Food security, drinking water, energy, schooling, basic healthcare, housing	**Personal resources for development:** Freely disposable time, income, skills, health	**Risk protection:** Disaster prevention, crime prevention, insurances	**Rights & participation:** Gender equality, empowerment, freedom of speech
	SDGs 6, 7, partly 2, 3, 4, 8, 11	SDGs 1, 3, 4, 5	SDGs 1, 2, 3, 5,	SDGs 5, 16
1.b) Levels of change (output-outcome-impact, or means-ends)	• *Access-to* is output, *usage of* is outcome • Basic services as such are rather means (setting conditions for development)	Resources as such are often an outcome resulting from basic services.	Better observed as means, requiring a clear theory of change. Ends hard to observe.	• Rather difficult to observe at all levels; theories of change tend to be complex.
2.a) Scale of measurement	• Often dichotomous scale (defined minimum service level reached or not).	• Often measurable on a continuous scale, such as hours or US$	• Different scales applicable.	• Different scales applicable.
2.b) Permanence	• Non-permanent (access to services can always lost)	• Partly permanent (e.g., skills), partly non-permanent (e.g., income)	• Preventive measures generally non-permanent.	• Successes not necessarily permanent.
2.c), 2.d) General quantifiability and suitable types of unit	**Mostly well quantifiable** as: • Person-years with access to a certain service level (output-level) • Person-years making use of certain service level (outcome-level)	**Often well quantifiable** as: • US$, living wages, or hours gained in total • Person-years with a defined minimum achievement (if non-permanent, e.g., gaining a defined extra-income) • Number of persons benefitting (if permanent, e.g., able to read)	**Partly quantifiable at output-level**, e.g., as person-years with minimum risk protection **Hardly quantifiable at outcome-level**, e.g., as reduced long-term damages)	**Hard to capture quantitively.** Often proxy-indicators used, often results for other targets specified for women) Example: Total sum of loans taken by women
3.a) Baseline or counterfactual	• Frequently defined as absence of minimum service level. • Often self-selection of beneficiaries or geographic selection.	• Often set for selected aspects only (e.g., time spent for water fetching, not regarding total domestic work). • Often varying between individuals	• Can be defined both at output level (such as health insurance coverage) and outcome level (such as certain mortalities)	• Baseline often highly complex and therefore difficult to establish.
3.b) Attribution	• Generally clear for physical solutions (e.g., water filters responsible for access to clean water)	• Often difficult due to individually varying baselines and multiple causalities (e.g., for health)	• Often problematic, mostly only attributable at output-level.	• Difficult since often addressed at policy-level.
3.c) Means of verification	• Often physical checks of service-providing facilities, together with surveys on usage.	• Often self-reporting, standardized tests on skills, modeling • Health, skills hard to verify at ends-level	• Rather verifiable at means-level. Outcome estimations by long-term evaluations.	• Often hard to verify, rather estimated by proxy-indicators
4. Quality control (typical issues)	Issues with dynamic baselines.	Courtesy bias in self-reporting, issues with baselines and attribution	Quantification at means-level is inaccurate	Generally difficult to quantify.

3.2 Nature and Environment: The categories defined below represent different approaches of targeting nature and environment, with common characteristics of measurability. The table may be used to orient the assessment of standardized indicators. → ***Detailed explanations of concepts are found in section 4!***

1.a) Category of SDG target referring to: Nature and Environment	**Specific threat mitigation**	**Protection of ecosystems**	**Direct management or manipulation of ecosystems**
	Tackling of land use change, invasive species, poaching, GHG, pollutants, …	Establishment of protected areas as generalized threat prevention, often combined with specific threat mitigation	Adaptation to climate change, forest management, fisheries, reforestation, re-introduction of species… Often directly effects human wellbeing, e.g., adaptation
	SDGs 12,13,14,15	SDGs 14,15	SDGs 13,14,15
1.b) Level of Change measured	A terminology of *means-ends* often fits better here than *output-outcome-impact*		
	• Means: Threat mitigation; • Ends: State of ecosystem	• Means: Protection status, management of protected area • Ends: State of ecosystem	• Means: Management or Manipulating activity • Ends: State of ecosystem or ecosystem service
2.a) Scale of measurement	Threats are often measured on continuous scales.	Protection status categorical/dichotomous	Varying scales applicable
	State of ecosystems measured in varying scales.		
2.b) Permanence	Positive change is generally non-permanent; there is always a possibility of ecosystems deteriorating and threats increasing.		
2.c), 2.d) General quantifiability and suitable types of unit	**Often well quantifiable** at means-level as: • tons of GHG/pollutants, … • ha of land use expansion, • population size of invasive species, …	**Fairly quantifiable** at means-level as: • Area of ecosystem under a certain protection status • Existence of management plans, …	**Partly well quantifiable** at means-level as: • Trees planted (vs. forest established), resources extracted, strategies developed, ha of ecosystem under sustainable management/certification
	Difficult to quantify at ecosystem level (ends-level): • Rather long-term observations of population sizes, ha of land reforested, species diversity, biodiversity indices, proxy indicators for ecosystem integrity/functionality, generalized categories of conservation state or endangerment		
3.a) Baseline / counterfactual	For threats, baselines are often business-as-usual scenarios (continuation of an existing trend), if assessed at all.	Systematic counterfactual analyses are rarely done for protected areas. (But can be meaningful for prioritizing areas).	Baselines are often business-as-usual scenarios, if assessed at all.
3.b) Attribution	Often difficult to attribute observed change (or maintenance of conservation status) to specific project activities. But often worth trying.		
3.c) Means of verification	Generally easier at means-level (direct observation of threat intensity, protection status or resource extraction). Often difficult for the condition of ecosystems or ecosystem services (long-term monitoring, proxy indicators, …). Link to threats not always clear.		
4. Quality control (typical issues)	Manyfold. For example, there is a risk of inaccuracies and deficiencies due to the complexity of interactions, such as neglection of ecosystem functionality. Threat assessment for conservation projects is often rather anecdotal.		

4 Explanation of concepts

Step 1: Analysis of the value to be measured

1a) Underlying intrinsic values, equivalence and categories of SDG targets

Human life

The value of a human life is widely accepted as an intrinsic value, as reflected, for example, in the universal declaration of human rights[1]. In medical evaluations, life years are often evaluated as equivalents[2]. Quantification is sometimes conducted by measuring years of life gained (ADALYs[1]), and there is also a concept of considering the quality of life by calculating so-called QALYs[3] that represent life years (gained by a specific therapy, for example), with a "utility" value referring to the quality of life (originally in relation to the health state).

Nature

While there is also wide recognition of an intrinsic value of nature, it is far more complex as a concept than human life[4]. It can be defined on different levels, ranging from genes over individuals and species to ecosystems and biomes.

Besides an intrinsic value, the value of nature is often defined by its value for humans, i.e., through ecosystem services. Such services are often equivalent within a defined context, but not globally (such as fresh water within a defined watershed).

Categories of SDG approaches

SDG targets are categorized here according to non-exclusive approaches of pursuing sustainable development. These approaches share common characteristics of measurability and can provide orientation for assessing SDG-related change (see section 3). There are some existing categorizations of SDGs; clearest is the differentiation between human- and nature-based categories, and the definition of "basic needs" or "essential needs"[5] within the former. The other categories applied here are inspired by existing classifications, but basically represent results or own reflections based on measurability.

Level of change

Change can be measured at different levels, and awareness about these levels of change can be very helpful for the design of indicators and monitoring schemes. A classical theory of change distinguishes *inputs* (resources used, such as bricks), *outputs* (direct results, such as toilets constructed), *outcomes* (second level effects, usage of toilets) and *impacts* (such as improved health conditions).[6]

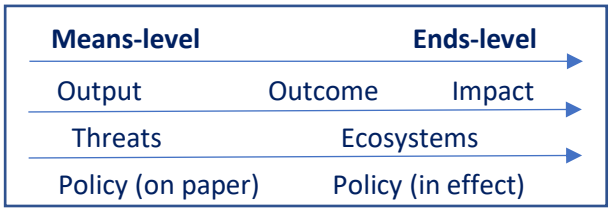
But often, exact assignment of levels is difficult. Therefore, the range between output and impact can also be seen as a spectrum ranging from the *means level* to the *ends level*. Moreover, the concept can be extended beyond the context of a specific project. SDG indicators do originally not refer to projects, but still may be assigned to a level of change. For example, SDG indicator 3.2.2 (Neonatal mortality

[1] www.un.org/en/universal-declaration-human-rights
[2] The global burden of disease, 1990–2020: www.nature.com/articles/nm1198_1241
[3] The use of QALYs in health care decision making: www.sciencedirect.com/science/article/abs/pii/0277953689900300
[4] The Society for Conservation Biology: https://conbio.onlinelibrary.wiley.com/doi/abs/10.1046/j.1523-1739.2000.98362.x
[5] Unravelling the complexity in achieving the 17 sustainable-development goals | National Science Review: https://academic.oup.com/nsr/article/6/3/386/5381567?login=true
[6] https://www.betterevaluation.org/sites/default/files/Define%20-%20Compact.pdf

rate) is rather an *ends level* indicator, while 3.1.2 (Proportion of births attended by skilled health personnel) is closer to the *means level*, also serving as a tool to achieve 3.2.2.

Projects targeting nature often target threats to ecosystems as a means to improve the state of ecosystems. Measuring such project's success may be most accurate if monitoring the development of the threats tackled, but only if there is a stringent theory of change showing the relation between threats and the state of ecosystems. Moreover, monitoring the ecosystems' state seems equally important.

Projects may contribute to the formulation of policies (e.g., municipal rules to prevent wildfires). Such policies can be both *means* (while existing on paper) and *ends* (if being effective as an institutionalization and appropriation of change). The concept of levels of change is thus understood in a very broad sense:



Some implications of assessing the level of impact:

- Results that are closer to the *means level* are generally easier to verify and quantify, and much easier to attribute (see below).
- Results closer to the *ends level* are often more meaningful, but often difficult to observe and only visible on the long term.

Change should be measured at different levels at the same time. However, depending on the project type, certain levels can be more meaningful than others. For example: The success of a renewable energy project in mitigating climate change can be quantified relatively close to the *means-level*, by metering the amount of renewable electricity fed into a grid, allowing to quantify GHG reduction. Measuring a direct impact on climate change at the *ends-level* is obviously impossible. But a project promoting literacy cannot be measured sensibly at the *means-level*, e.g., by the number of lessons told; meaningful results require also direct measurement of the improved literacy skills of the target group, closer to the *ends-level*.

Step 2: Definition of a standardized unit

If project results are to be comparable between projects, they need to be measured in standardized units. Ideally, such units are perfectly equivalent, thereby allowing for unlimited comparisons in space and time. Such equivalence is hardly reached. An example coming close is a ton of CO_2e emitted which can be deemed globally fungible: It is assumed to have the same effect on the climate, no matter where on the globe it is emitted or reduced; and it is also assumed to be temporally equivalent, since the moment of emission/reduction is (relatively) irrelevant.

2a) Scale

Most original SDG indicators apply a relative scale, for example, measuring proportions of a population. But a project's contribution to SDG achievement needs to be summable and measured in absolute numbers, in order to allow for comparing different projects with different target group sizes. Proportions can often be converted to absolute numbers (e.g., indicating "number of people" instead of "percentage of the population" living below the poverty line). But in many cases, original indicators need to be transformed substantially for a meaningful representation of project results. (And some

original SDG indicators do not make sense at the project-level, if, for example, referring to national budgets.)

Measurements of progress concerning people is often based on dichotomous (yes/no) values. Examples include life below the poverty line or access to a defined minimum level of basic services (such as basic access to drinking water[7]). Dichotomy is hereby often derived by simplifying categories (e.g., grouping levels below and above "basic access"[6]) or by setting thresholds for continuous scales (such as a poverty line based on income).

Also, continuous scales as such are often suitable for project-level SDG measurement, if presenting results as accumulated value achieved, such as total additional income generated, or total hectares of forest planted.

Threats to ecosystems can often be measured at continuous scales, the state of ecosystems or endangered species is often approximated by using categorical scales.

2 b) Permanence

Progress can be automatically durable. Skills like the ability to swim are normally learned for a lifetime, and the benefit of birth assistance avoiding maternal death is timeless. But in many cases, progress is not permanent. The access to health services or electricity may be lost at any time. The quantified value of a project's results needs to consider permanence. Even an achievement that is lost after some time may be assigned a quantified value. If a project providing efficient wood stoves stops operating after some years (e.g., if efficient cookstoves are replaced by government-sponsored LPG cookers), no further GHG savings can be claimed by the project. However, the accumulated amount of emissions reduced from unsustainably harvested firewood till then maintains its value. The same principle can be applied to aspects of human life: A number of life-years with a certain quality represents a quantifiable value, in analogy to QALYs[2] (see above), even if the quality is then lost. Even equivalence between persons may be defined: The value of two persons having access to electricity for one year may be deemed equal to the value of one person having electricity for two years.

Nature-based SDG targets are generally non-permanent, and even temporary achievements cannot be standardized as shown above. The survival of a threatened species for some additional years probably has a neglectable value, but in special cases, some years gained for the survival of some individuals may finally guarantee the survival of the entire species, thereby gaining a very high value.

2 c) Assign a standardized unit

Three types of units have been identified that allow for quantitative measurements:

- **Person-years** with a defined minimum achievement measure non-permanent, dichotomous results (such as person-years with basic drinking water access). Person-years are defined as equivalents: Two persons with access for one year = one person with access for two years.
- **Number of persons** with a defined minimum achievement apply to permanent, dichotomous results, such as people vaccinated or people able to read
- Total **US$** earned, **hours** gained or **ha reforested** apply to continuous results and can totalize a project's achievements. It may be necessary to complement the analysis with a measurement considering the distribution of benefits among the target population, such as person-years with minimum gainings.

2.d) General quantifiability

[7] https://data.unicef.org/resources/progress-drinking-water-sanitation-hygiene-2017-update-sdg-baselines/

A systematic assessment of the criteria explained above may be helpful to determine the quantifiability of certain aspects of sustainable development. The evaluation of categories in section 3 provides further guidance in this regard. It is however important to keep in mind that quantifiability does not tell anything about the importance of results.

Step 3: Measurement of change

3a) Baseline scenario or counterfactual

A baseline scenario or counterfactual defines a hypothetical non-project case against which a project's results are measured. It cannot be observed directly. Two features are important for the quality and accuracy of a baseline scenario:

- Independence: Ideally, an independent control group is determined randomly, preferably before the project start and from the same population as targeted, and then monitored in parallel with the project[8]. Establishing independent control groups is a standard in empirical natural science (like medical studies), but rarely found in projects related to sustainable development. Here, it is more common to determine baselines by assessing the target group itself before the project starts. Average income or health characteristics may be determined in specific surveys or simply by applying available official statistics. Or baselines may be established implicitly by simply selecting the target group according to certain baseline criteria, e.g., by geographical selection (a project area without electricity connection may be defined for an electrification project), or by self-selection (only beneficiaries without access to drinking water are probably interested in buying a water filter). Among others, such selective baselines are common in RBF programmes dealing with basic services. Baselines that are fixed ex-ante do however disregard external developments after the project start. Therefore, a good baseline scenario should also be:
- Dynamic: A changing baseline may alter a project's results, if, for example, a government extends the electricity grid to former offgrid areas where a project installed isolated mini grids. An independent control group monitored in parallel automatically accounts for external factors after the project start. Where the baseline is established by assessing the target group itself, it can be adjusted regularly according to available official statistics such as economic growth. Sometimes, ex-ante baselines are modelled to account for dynamic future development, e.g., by modelling expected deforestation in the project area of a REDD+ project.

In projects targeting ecosystems, the practice of establishing counterfactual thinking less common, probably because results are more difficult to observe and counterfactuals more difficult to establish[9,10].

A counterfactual is generally easier to determine if focusing on isolated effects pursued by a project, such as reduced expenses on firewood due to usage of an efficient cookstoves, instead of conducting an overall assessment of the beneficiaries' income – which may however be more meaningful.

3b) Attribution

A comparison of results with an independent control group can reveal the share of observed change attributable to a project; in this case, the question about attribution is answered together with counterfactual comparison. In GHG reduction projects, a project is generally regarded "additional" or "non-additional", i.e., either all or none of the change observed is attributed to it; often by arguing

[8] White, H., & Raitzer, D. A. (2017). *Impact evaluation of development interventions: A practical guide*. Asian Development Bank.
[9] Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New directions for evaluation*, *2009*(122), 75-84.
[10] Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., ... & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*(1), 58-64.

that a private project becomes only profitable with carbon funding – which represents a simplification of attribution assessment.

Attribution may become an issue if in a project area, similar activities are funded by different agents; e.g., a donor funding a pilot project introducing a new technology might claim a large share of future achievements, but so may donors do who funded an upscaling of the activity. Carbon markets refer to this problem as "double-counting".

Attribution is generally easier to show for results achieved soon after project start and for results closer to the *means-level* (therefore, RBF schemes use rather output indicators). Attribution is particularly challenging for projects targeting nature (Ferraro 2009).

3c) Means of change verification

Possible ways of practical change verification are highly dependent on the type of change assessed. Some non-exclusive examples for means of verification include:

- Direct observation of physical change (such as drinking water available, trees planted, individuals of threatened species found, forest extension by remote sensing…)
- Surveys based on interviews (such as household income, habits of personnel hygiene, food insecurity (risk of framing, courtesy bias in self-reporting)
- Use of standard registries and databases (such as number of patients attended, school attendance, employment)
- Standardized tests (such as literacy)
- Observation by proxy-indicators (such as indicator species for biodiversity)
- Statistical modelling (such as ADALYs)

These means may be combined with approaches such as random sampling, plausibility checks and third-party auditing.

Results that are theoretically well quantifiable may be difficult to observe (e.g., ADALYs). Vice-versa, there are results relatively easy to observe (e.g., forest cover change) but hard to quantify as a result of a certain project. As in the case of attribution, verification is generally easier for results closer to the *means-level*.

Step 4: Quality Control

4a) Biases and shortcomings

There are typical biases occurring during the verification of change, including:

- Selection bias: For example, self-selected participants in a training program may have higher motivation and skills than the average populations, leading to a biased baseline.
- Courtesy bias: In surveys based on interviews, interviewees may report positive results just to please the person interviewing them.
- Framing: Multiple choice questionnaires may omit important answer options and thereby lead to biased results.
- Confusion bias: For example, a forest claimed to be conserved by a REDD+ project may persist due to inaccessibility rather than the project's activities.
- Leakage (overspill): Negative influencing factors are pushed outside the project boundary, leading, for example, to a shift of deforestation)
- Shortcomings in indicator definition, if, for example, access to health services is defined by distance to facilities, without considering a river that needs to be crossed.
-

4b) Risks and unintended effects

Besides intended change, projects can also cause unintended and/or unwanted change or imply new risks. While intended change (positive effects) can be assessed specifically, an observation of possible negative effects generally requires a systematic screening of pre-defined potential risks (do-no-harm assessment, safeguarding principles etc., REF GS). Such a screening is an essential part of change assessment, but not subject of this document.

4c) Reliability of quantified results

Quantified results may also be helpful if their accuracy and reliability is not optimal (due to low quantifiability in general, accuracy of available data etc.). It is suggested to communicate the reliability in fixed categories, with additional explanations.

- *Verified*: Quantified results have been checked externally, and/or they are verifiable by publicly accessible means (such as Google Earth showing reforested areas).
- *Estimated*: Some verification occurred, but quantification relies on estimates with important uncertainties. Results are made plausible by cross-checks etc.
- *Narrative*: Results rely on generalized statements of beneficiaries, and/or are generally hard to quantify. Still it is recommendable to show the plausibility by cross-checks and in-depth interviews with selected users, for example.

4d) Non-quantifiable achievements

Results may be of limited quantifiability by nature, but still represent the most important achievements of a project (e.g., if targeting human rights). Other ways must be found to communicate such results; and striving for quantified results should not lead to a neglection of unquantifiable results.

5 Conclusions

Official SDG indicators are often hard to apply at the project-level, among others, because they often apply relative metrics which do not allow for comparisons between projects. But with some adjustments, many SDG achievements can be measured in universally comparable and standardized units, such as person-years.

However, such a standardization is always just an approximation to the reality, full equivalence between different project's achievements can probably never be reached. Therefore, the creation of tradable units such as carbon credits is probably not advisable in most cases.

It is found that human wellbeing can be easier to standardize than results related to nature and environment, due to the universal value of human life. Access to basic services can be standardized quite well, while the promotion of rights and participation can hardly be captured in a standardized way. Nature is mostly targeted indirectly by projects though the mitigation of threats. Observing a project's effects on the state of ecosystems is therefore often difficult; therefore, measuring effects on threats may be equally or even more important, together with the establishment of a stringent theory of change. Counterfactual thinking (assessing would have happened without the project) is deemed important in any case.

But even where standardization is difficult, as systematic application of the suggested approach may increase the understanding and the quality of measuring a project's results. One important conclusion is that quantified achievement should always consider the duration of effects over time. Quantification should therefore occur ex-post, based on regular monitoring.

Standardization of project results based on systematic assessment can be helpful in several ways, as long as one keeps in mind that it will never capture the entire breadth of effects reached by projects.

Annex 1 Examples of selected project-level SDG indicators in practice

SDG 3: Health impacts of kitchen smoke

Solar Cookers disseminated in a project of Fair Climate Fund allow refugees in Chad to replace an important part of cooking on traditional wood stoves[11]. This can avoid smoke exposure and thereby avert negative health impacts (mainly respiratory diseases). Measuring this effect is easier by measuring smoke exposure, while direct measurement of health impact is challenging. Indicators for both aspects are defined here.

Project-level SDG indicator	**Reduced kitchen smoke exposure (output/means-level)** Derived from original SDG indicator 3.9.1: Mortality rate attributed to household and ambient air pollution
Analysis of the value to be measured	**Category:** Risk prevention, also related to basic services. **Level of change:** Represents an output (means) for the outcome (ends) of health improvements, better measurable than the latter.
Standardized unit	Person-years of avoided kitchen smoke exposure. Example: 3,000 person-years resulting from 1,000 solar cookers in use over 5 years, with an average usage rate of 60% (40% of cooking still done with baseline stoves). One person is assumed to do the cooking. If different persons replace each other, this does not affect the calculation of person-years.
Baseline / Counterfactual	The baseline consists in cooking on traditional wood stoves. However, details of smoke concentration and ventilation are not assessed, therefore, the baseline simply constitutes an exposure to “typical smoke concentrations of traditional wood stoves”.
Attribution	The achievement is clearly attributable to the project distributing solar cookers.
Means of verification	Physical checks of the functionality of solar cookers, together with user surveys on usage patterns, some plausibility checks.
Quality control	Risk of courtesy bias in user survey, particularly regarding the continuous use of baseline stoves. Inaccuracy due to generalized baseline (just typical smoke exposure).
Conclusion	The indicator is relatively well quantifiable. Higher accuracy can be achieved by assessing more details of baseline smoke exposure (smoke concentration, exposure times, occasional outdoor cooking, etc.). It is recommendable to assess health outcomes in parallel.

Project-level SDG indicator	**Reduced health impacts of kitchen smoke (outcome/ends-level)** Derived from original SDG indicator 3.9.1: Mortality rate attributed to household and ambient air pollution
Analysis of the value to be measured	**Category:** Personal resources for development **Level of change:** Represents an outcome (ends), caused by the reduced smoke exposure (output or means).

[11] www.fairclimatefund.nl/en/projects/chad-solar-cookers-for-refugee-families

Standardized unit	a) Perception of health impacts: Can only represent a qualitative result, since it cannot be assessed how much improvement a person perceives. b) ADALYs: Quantifies life years saved, but only based on statistical modeling, which is based on measuring the means (reduced smoke exposure).
Baseline / Counterfactual	When asking about perceived health improvements, the baseline is not assessed directly. When calculating ADALYs, the baseline is given trough statistics based on measuring the means (reduced smoke exposure).
Attribution	Attribution is not assessed specifically.
Means of verification	Interviews for perceived impacts. For ADALYs, only smoke exposure is verified. Direct verification of health outcomes is possible, but requires thorough long-term studies.
Quality control	When asking for perceived health effects, there is a risk of courtesy biases and framing which may lead to over-estimation. Results are inaccurate and can only be used as qualitative or narrative findings. For ADALYs, reliability depends on the quality of measuring smoke exposure.
Conclusion	The indicator can basically add some qualitative and illustrative information to results obtained when measuring smoke exposure.

SDG 6: Access to drinking water

The Dutch NGO Aqua for All promotes the development of Household Water Treatment and Safe Storage (HWTS) market in Ethiopia and Malawi. Through public and private partners, filters are sold with a small profit margin. Aqua for All covers initial investment costs, links private and public sector, provides necessary trainings, and does advocacy with relevant government institutions.

Filters allow households to use and purify water from any source, not necessarily having to walk for long distances to water kiosks or public fountains. This often allows users to gain basic access to drinking water. (Basic service level according to WHO definitions[12]: Drinking water from an improved source, with collection up to 30 minutes for a round trip, including queuing. Improved source: piped water, boreholes or tubewells, protected dug wells, protected springs, and packaged or delivered water.)

Project-level SDG indicator	**Improved access to drinking water** Derived from original SDG indicator 6.1.1: Proportion of population using safely managed drinking water services
Analysis of the value to be measured	Category: Basic services Level of change: The possession of water filters can be regarded an output. The relevant outcome for the indicator under discussion is usage of these filters in replacement of a sub-basic drinking water level.
Standardized unit	Person-years of making use of a basic service quality of drinking water Example: 2000 Person-years (400 persons gaining access during 5 years) Basic service level according to: https://data.unicef.org/resources/progress-drinking-water-sanitation-hygiene-2017-update-sdg-baselines/

[12] https://data.unicef.org/resources/progress-drinking-water-sanitation-hygiene-2017-update-sdg-baselines

Baseline / Counterfactual	Selective baseline within a geographical focus (users decide themselves if they buy a filter; filters are offered in areas with limited drinking water access). It is therefore possible that filters are used by households that already have a basic access to drinking water. This needs to be assessed for an accurate quantification of the indicator. Moreover, the baseline needs to be re-assessed regularly. For example, new water kiosks may open in less distance, implying a basic service level even in the baseline scenario.
Attribution	The achievement is clearly attributable to the project distributing water filters.
Means of verification	Physical checks of the functionality of the technology, user surveys on usage patterns, including re-assessment of the baseline.
Quality control	Risk of courtesy bias in user survey. Risk of changing baseline during project operation. Risk of filter distribution disincentivizing public drinking water programs.
Conclusion	The indicator is relatively well quantifiable. Comparisons between similar projects may be limited by possible variations in the baseline scenario – some projects may attend households with service levels closer to "basic" than others, the latter would then achieve a greater approvement than the former, which would not be observed by the indicator. It is recommendable to assess other indicators such as time savings in parallel.

SDG 5: Saving time from domestic work

See above (Aqua for All)

Project-level SDG indicator	**Time saved from drinking water collection** Derived from original SDG indicator 5.4.1 (Proportion of time spent on unpaid domestic and care work)
Analysis of the value to be measured	Category: Personal resources for development Level of change: The possession of water filters can be regarded an output; time savings in consequence of using these filters is an outcome.
Standardized unit	Total hours saved for the target group 365,000 hours of freely disposable time13 (400 persons gaining average 30 min hour daily during 5 years) Additionally, it is recommendable to include a measure for the distribution of time savings within the target group.
Baseline / Counterfactual	The baseline varies for different households, depending on the distance of the next drinking water source, average queuing time, but also average water consumption. It may be approximated jointly for neighborhoods with similar conditions. The baseline needs to be re-assessed regularly. For example, new water kiosks may open in less distance, implying reduced water collection time.
Attribution	The achievement is clearly attributable to the project distributing water filters.
Means of verification	Physical checks of the functionality of the technology, user surveys on usage patterns, including re-assessment of the baseline.

[13] Hobbes, M., De Groot, W. T., Van Der Voet, E., & Sarkhel, S. (2011). Freely disposable time: A time and money integrated measure of poverty and freedom. *World Development*, *39*(12), 2055-2068.

Quality control	Risk of courtesy bias in user survey. Risk of changing baseline during project operation. Risk of disregarding a social value of the time in cue for water fetching.
Conclusion	The indicator is fairly well quantifiable. Accurate estimates of time savings may however be jeopardized by varying baseline conditions. Time required for water fetching may change during the project, particularly queuing. This may also limit the possibilities to compare similar. It is recommendable to assess other indicators such as access to drinking water in parallel.

SDG 15: Meadow breeding birds in farm land

Under the EU's Common Agricultural Policy, conservation of meadow breeding birds is generally promoted by compensating farmers who implement measures considered favorable for these birds (e.g., reduced and adjusted mowing). In an alternative, result-based approach, farmers only receive incentives if successful events of breeding take place at their farms, independently of the measures applied.

Project-level SDG indicator	**Meadow breeders in farmland** Derived from original SDG indicator 15.5.1 Red List Index
Analysis of the value to be measured	Category: Ecosystem management (together with direct threat mitigation) Level of change: Means: Offering suitable conditions for breeding Ends: Breeding events occurring, and finally, improved viability of the entire population
Standardized unit	Breeding events (accumulated over the years) (Example: 80 breeding events (20 sites, in average 4 breeding events in 10 years of project duration…)
Baseline / Counterfactual	At the farm level, absence of birds is assumed, not assessed in particular. At the level of the project in total, the baseline constitutes breeding events in former years (independent control group virtually impossible to establish).
Attribution	Breeding events can be attributed to offering suitable conditions to birds, but absence of birds can happen in spite of suitable conditions present.
Means of verification	Direct observation by ornithologists or trained farmers, documented by pictures
Quality control	Due to issues with attribution, it is difficult to compare the results of different attempts to enable breedings. But still, measuring breedings is relatively simple and clearly indicates success.
Conclusions	Result-bases programs will probably work best where in the past, breedings were common. Farmers will probably be familiar with the measures to be taken, and the chance of being successful is probably high enough as a motivation. But if new breedings shall be enabled where these have rarely been observed in the past, it may work better to directly fund measures, probably with a bonus for success.

Annex 2: List of potential project-level indicators per SDG

SDG and comment	Generalized indicator and analysis of value measured (in brackets: original SDG indicator referred to)	Notes on practical change measurement
1. End poverty in all its forms everywhere *Poverty is multi-dimensional. Some aspects (such as access to basic services) are better covered under other SDGs.*	**Available personal/household budget (1.2.1):** *Total additional income (or reduced expenditures) of target group, in US$ and/or equivalents of national living wages* (see *wageindicator.org, for example*) May be good to include an indicator on distribution of additional budget among the target group. Level of change: Can be regarded output if income originating from project budget, outcome if generated from for activities stimulated by the project.	Easier to assess for isolated, project-specific aspects such as firewood expenses in an efficient cookstove project. An assessment of the total budget is more challenging (particularly for attribution), but also more meaningful. Important to consider dynamic baselines (such as changing fuelwood prices). Ideally, based on calculations using external data (such as fuelwood prices), cross-checked by self-reporting.
	For most other aspects, standardized quantification is difficult at project level (social protection, disaster losses, secure land tenure…).	
2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture. *Meaningful results require long-term assessments.*	**Food security (2.1.2):** *Person-years with: FIES (Food Insecurity Experience Scale) reduced by a certain level (for example, from moderate or worse to mild or better)* Level of change: Is more an output if food provided by the project, more an outcome if self-produced in consequence of the project.	Attribution is difficult, ideally based on independent control groups. Verification based on a standardized questionnaire with subjective questions - risk of courtesy bias particularly for projects.
	Productivity of small-scale agriculture (2.3.1): Total increase of smallholder production of a certain crop among the target population. Important to include an indicator on distribution of additional budget among the target group. Is rather an outcome. Important to consider long-term sustainability of measures (simple increase of mineral fertilizer may not me sustainable).	Attribution difficult, ideally based on independent control groups. Complex relations between yields of certain products, climatic variations, income from sales and food security. Needs integrated assessment of many different factors.
	For most other aspects, standardized quantification is difficult at project level (reduced undernourishment and stunting, agricultural prices, genetic resources, …). Proxy-indicators may be defined for these aspects.	
3. Ensure healthy lives and promote well-being for all at all ages *Quantification easier for aspects that are output/means such as access to services or exposure to health-risks.*	**Access to basic health services (3.8.1, 3.b.1):** Person-years with: Access to and usage of a defined minimum health service (such as medical center within 1 hour travel distance). Number of persons with: Assisted births, specific vaccinations. Access-to is output, usage-of and the effect-of is outcome.	Baseline definition can be by geographic selection (unattended areas). Focus on a few selected health-services may cause neglection of others. Attribution difficult for health effects.
	Access to family planning (3.7.1): *Similar to Access to basic health services*	
	Healthy personal environment (3.9.1, 3.9.2, 3.9.3): Person-years with: Exposure to harmful conditions reduced by a certain level (indoor air pollution, drinking water, …) Need to define thresholds (smoke concentration, water contamination, …) Exposure-to is more an output (means) indicator.	Much easier to verify than effects. Baseline-setting (threshold definition) can be challenging. Important to consider dynamic baselines (such as changing access to drinking water in the target area).
	Health effects (3.9.1, 3.9.2, 3.9.3) Person-years with reduced symptoms like cough or diarrhea, or, eventually, ADALYs (life-years gained as such). Can regarded outcome or even impact.	Hard to verify. Baseline and attribution very challenging; normally requiring long-term studies and independent control groups. ADALYs calculated statistically, generally not verified empirically.
	For most outcome-level health aspects, standardized quantification difficult at project level (such as child mortality, incidence of diseases. Access to health insurance may be quantifiable, if jobs are created with social benefits).	
4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all *Quantification easier for output/means (attendance). Skill mostly difficult to measure and attribute.*	**Attendance to schools/other education offers (4.1.1, 4.2.2, 4.3.1):** Person-years with minimum attendance to schools/ other education offers. Clearly an output	A clear definition of minimum attendance may be difficult, particularly if considering the quality of education. Results may be difficult to attribute to a specific project.
	Number of persons who gained certain skills (literacy, numeracy) (4.1.1, 4.4.1, 4.6.1). Outcome of schooling/education.	Verified by standardized test which may be challenging; control groups needed for quantification and attribution of progress.
	Parity of attendance to education (4.5.1) may be evaluated by defining changing person-years for specific groups (male/female, urban/rural, …). Health/learning/psychosocial development of children (4.2.1) difficult to evaluate as a result of a specific project.	
Goal 5. Achieve gender equality and empower all women and girls *Results often difficult to quantify and attribute*	**Freely disposable time (5.4.1):** Total accumulated freely disposable time generated for the target group. Originally applying to reduced domestic work by women, but may be applicable in a wider context. Can generally be regarded an outcome.	Easier to assess for isolated, project-specific aspects such as time saved from water fetching, instead of assessing total disposable time (which is challenging). Baseline may be individually different.

to a specific project, particularly if referring to women's rights.		Time savings may imply a loss of social value (if, for example, cuing for water allows meeting friends).
	Relative measures such as share of women in certain positions etc. make sense within the context of a specific project and may be difficult to transform to absolute measures. This can be tried from case to case, for example, by measuring the total turnover managed by women in managerial positions. But in many cases, the complexity of SDG 5 is difficult to standardize in a quantitative way.	
6. Ensure availability and sustainable management of water and sanitation for all *Often quantifiable as access to services. Where referring to water resources, only comparable within specific boundaries such as watersheds.*	**Access to drinking water (6.1.1):** Person-years with: Access to and usage of basic service quality of drinking water, e.g. based on WHO definitions (threshold dependent) Access-to is output, usage-of is outcome.	Baseline conditions may differ for each household (e.g., time to fetch water). Verification by checking operation and usage of physical appliances such as water filters or well.
	Access to toilets and hygiene (6.2.1): Person-years with: Access to and usage of toilets and personal hygiene facilities Access-to is output, usage-of is outcome.	Important to distinguish between public and personal facilities.
	Possible indicators with limited quantifiability: - Within defined urban contexts: Access to and use of waste collection and wastewater treatment in urban areas (6.2.1), measurable in person-years. - Within defined boundaries such as watersheds: Efficiency of freshwater usage (e.g. for irrigation) (6.4.1), measurable e.g. in liters/yield, percentage of freshwater withdrawal from available resources (6.4.2) - Households in settlements with local management of water resources (6.b.1) (in person-years)	
7. Ensure access to affordable, reliable, sustainable and modern energy for all *Possible additional indicator: Access to clean lighting.*	**Household electrification (7.1.1):** Person-years with: Access to and usage of defined minimum service of grid electricity. Needs threshold-definition (for example, regular availability of power). Access-to is output, usage-of is outcome.	Verification by physical checks. Need to observe government-driven grid extension (dynamic baseline).
	Reliance on clean cooking fuels (7.1.2): *Person-years with: Access to and usage of clean cooking fuel (biogas, alcohol from renewable sources…)* Original SDG indicator refers to clean fuels. Also reduced consumption due to increased efficiency may also be quantified. Access-to is output, usage-of is outcome.	Verification by physical checks of operation and usage of cooking appliance and/or fuel availability.
	Possible indicators with limited quantifiability, within the boundary of a grid (7.2.1): - Renewable energy share in the project area's energy production; kWh saved due to energy efficiency.	
8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work *Results are often difficult to isolate from the general economic context.*	**Job openings (8.3.1, 8.5.2**) *Person-years of employment generated* Can be disaggregated by job quality, full vs. part time, social insurance etc. But job quality is difficult to standardize. Output if salary paid by project, outcome if from external sources.	Baseline assessment needs to consider if jobs are new and do not replace existing jobs. May be difficult to compare different types and qualities of jobs; and also, to define which jobs are created due to a specific project.
	Youth employment / education (8.6.1): *Similar to job openings.*	
	Earnings from work (8.5.1) *Total job earnings, in US$ and/or equivalents of national living wages* (see *wageindicator.org, for example*) May be good to consider also the distribution of earnings among the target group and the level of hourly payments. Can be regarded output if earnings are paid by the budget, outcome if generated from activities stimulated by the project.	Baseline assessment needs to consider if jobs are new and do not replace existing jobs.
	Total economic value created (8.1.1) Total US$ or living wages generated Output if measuring value created by direct project activities, outcome if considering alone-standing activities motivated by the project.	Calculations often based on assumptions difficult to verify. Attribution also challenging.
	GDP-related indicators are difficult to quantify as project results, as well as effects on child labour and workers's rights.	
9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	Often of limited relevance for project activities, and difficult to quantify in general.	
10. Reduce inequality within and among countries	Often difficult to measure in absolute numbers, hard standardize.	
11. Make cities and human settlements inclusive, safe, resilient and sustainable	Often difficult to standardize. of limited relevance for project activities. It may be possible to quantify standardized achievements with regard to basic housing, for example, or access to waste collection.	

Goal 12. Ensure sustainable consumption and production patterns	Probably the aspect of **domestic waste reduction (12.2.2)** is best measurable as project result, for example, the total amount of plastic waste avoided by a project.	
13. Take urgent action to combat climate change and its impacts *Original indicators need to be transformed substantially to be suitable at project-level.*	**Climate change mitigation (13.2.1)** Tons of GHG avoided or sequestered. Means-level, but very well linked to ends level (climate change) Threat avoided ToC clear Projects introducing climate-friendly technology: development pathways Climate-smart strategies (as mentioned in original SDG indicators) are essential but hard to quantify.	Many methodologies available on GHG avoidance by several carbon standards. Attribution (additionality) is often an issue, permanence with forestry projects and baseline definition with REDD+ projects.
	Adaptation (13.1.2, 13.2.1): Can be quantified as output/means (such as hectares of drought-resistant crops established), provided there is a stringent theory of change. Outcomes such as number of deaths avoided (13.1.2) or reduced damages caused by climate change are hard or impossible t to observe and attribute.	
	Other aspects included in original SDG indicators such as credibly quantified future pathways of GHG reduction and green technology adoption may be used as qualitative indicators by project.	
14. & 15 (similar in terms of measurement) 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development 15: Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss *Potential indicators are put in the order to categories defined in section 3.2, jointly for SDGs 14 and 15.*	**Direct threat mitigation** Tackling of land use change, invasive species, poaching, GHG, pollutants, … Measuring threats belongs to the means-level. Measuring the ends-level (effects on ecosystems) generally requires long-term evaluations of population sizes, proxy-indicators for ecosystem integrity/functionality, … - Marine plastic pollution / costal eutrophication (14.1.1) - Total amount of plastic waste / fertilizer leakage avoided at the source. - Hectares of land use expansion reduced. - Population size of invasive species - Individuals of a threatened species lost by poaching, etc.	The stringency of the theory of change decides over the quality of results (*how are ecosystems influenced by threats?*) Baseline and attribution are often problematic (representing a major issue with REDD+ projects, for example). Permanence is always an issue, implying the risk of positive achievements to be lost again.
	Protection of ecosystems and/ or species Establishment of protected areas as generalized threat prevention. Protection represents a measure at the means level. Measuring the ends-level (effects on ecosystems) generally requires long-term evaluations of population sizes, proxy-indicators for ecosystem integrity/functionality, … - Surface of protected areas established of a certain ecosystem type (14.5.1, 15.1.2) - Existence of management plans, funding, …	The simple existence of a protection status may not tell much about the effectiveness of protection. Systematic counterfactual analyses (what would happen without) are challenging are therefore rarely conducted for protected areas. But they can be meaningful, for example, when prioritizing areas.
	Direct management or manipulation of ecosystems Fisheries, forest management, reforestation, re-introduction of species… Often also measurable at the ends-level (forest area planted). - Size of populations of defined fish species in selected habitats (14.4.1, 14.7.1). - Total value of fish caught stemming from sustainable fishing (14.7.1) - Area of forest under sustainable forest management (means-level) (15.2.1) - Number of trees planted (means-level) - Area of forest planted (ends-level) - Population size of re-introduced species	It is often difficult to understand the full extent of effects of management or manipulation on ecosystems, due to the complexity of interactions.
16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	Most aspects are difficult to quantify at project level.	