Table of Contents

1. Preamble ................................................................. 3

2. Scope ........................................................................ 3

3. Audience .................................................................. 4

4. Definitions .................................................................. 4

5. Purpose of Benchmarking .......................................... 4

6. Typology ..................................................................... 7

7. Principles .................................................................... 10

8. Benchmarking Development Process ............................. 12

  8.1 Determine the purpose of the benchmark and who it is for ................................................. 12
    8.1.1 Audience ........................................................................ 12
    8.1.2 Purpose ........................................................................... 12

  8.2 Determine who or what is being benchmarked ................................................................. 12
    8.2.1 Type of entity ................................................................... 13
    8.2.2 Supply chain scope ............................................................ 13
    8.2.3 Market presence .............................................................. 14
8.3 Determine who will manage the benchmarking process ........................................ 14
8.4 Determine the content of the benchmark ................................................................. 15
  8.4.1 Considerations for developing benchmark content ................................................ 15
  8.4.2 Core elements for benchmarking standards systems .............................................. 16
  8.4.3 Alignment with international norms and guides ....................................................... 18
  8.4.4 Accommodating diverse approaches in a benchmark ........................................... 18
8.5 Determine the evaluation structure of the benchmark ................................................ 21
8.6 Determine the benchmarking implementation process ............................................... 21
  8.6.1 Steps in the benchmarking implementation process ................................................ 21
  8.6.2 Monitoring of continued alignment ..................................................................... 24
8.7 Determine how the results will be communicated ....................................................... 24
8.8 Further considerations ................................................................................................ 25
  8.8.1 Competence .......................................................................................................... 25
  8.8.2 Cost and complexity ............................................................................................ 26
  8.8.3 Improving the benchmark and process ................................................................ 26
  8.8.4 Government use of benchmarks .......................................................................... 27

Annex 1: Summary of Sustainability Benchmarking Programmes ........ 28

Annex 2: Benchmarking Good Practice Checklist ................................. 30


The following symbols are used throughout this guidance document:

- Action
- Take Note
- Idea
1. Preamble

The rapid rise of environmental crises and social upheaval, coupled with ambitious global commitments like the UN Sustainable Development Goals (SDGs), has pushed sustainability into the mainstream. Companies, governments and civil society are seeking tools they can use to evaluate sustainability performance and to recognise and reward good practice. For stakeholders wondering which tools to use or approaches to take, the landscape can seem bewildering and challenging to navigate. One response to this challenge has been the development of benchmarks to evaluate, compare and qualify sustainability tools and company performance.

A benchmark is a reference point against which something is evaluated. Sustainability benchmarks can evaluate a wide variety of entities, from the sustainability performance of companies to sustainability standards and certification. Benchmarking programmes define a specific reference point and carry out evaluations of sustainability policies, practices and tools against it. In this way, benchmarking programmes chart a path through the wilderness, providing users with comparable information about the benchmarked entities that then allows those users to choose between them. By way of example, governments can use benchmarking programmes to qualify sustainability standards for use in public procurement or for fulfilling company due diligence obligations; companies can identify which standards or sustainability tools are appropriate for their production needs; and NGOs can rank the sustainability performance of companies, creating an incentive for improvement.

The challenge with benchmarking is that there has been little guidance on how to develop and implement a credible benchmarking programme or exercises, leading to a proliferation of efforts of varying rigour, transparency and effectiveness. This is significant because it means that these programmes have the potential to recognise and reward lower performers, which limits the effectiveness of our collective response to today’s sustainability challenges.

Through this guidance, ISEAL aims to contribute a framework and practical set of good practices and considerations for those organisations and initiatives considering whether to carry out a benchmarking exercise or develop a benchmarking programme. We also aim to publicise existing benchmarking programmes as a means to limit the proliferation of benchmarks. Our goal is to support increased consistency and strengthening of sustainability benchmarking programmes so that they can effectively support better practices and a faster transition to a more sustainable world.

2. Scope

This guidance applies to any ongoing benchmarking programme for analysing or evaluating sustainability initiatives or performance. It can also be applied to one-off benchmarking exercises. While the guidance was developed in the context of benchmarking of voluntary sustainability standards (VSS), most of the guidance applies equally to benchmarking of sustainability policies or performance of different entities, e.g. companies, governments, NGOs or other stakeholders. Where there are differences in how these benchmarks are carried out, these are noted in the guidance.
The guidance does not propose criteria or requirements to be used in a benchmark but sets out considerations for developing those requirements.

3. Audience

The primary audience for this benchmarking guidance is the organisations and initiatives that are setting up or carrying out benchmarking exercises or programmes. The guidance can be used as a reference tool for these initiatives.

The guidance is also applicable for users of benchmarks or the entities that are benchmarked, to better understand what information they should be looking for or questions they should be asking from a benchmarking programme.

4. Definitions

Benchmark

- A benchmark is the reference point against which something is evaluated (noun)
- To benchmark is the act of determining (or judging) alignment with the fixed reference point (verb)

Benchmarking exercise

- A ‘one-off’ benchmark, which is not part of a long-term programme, policy or strategy. For example, a scoping report delivered by a consultant to inform a policy process might include a benchmarking exercise but does not constitute a benchmarking programme.

Benchmarking programme

- A structured and systematic way of carrying out evaluations against benchmarks, often coupled to specific organisational or policy goals.

Comparison

- A provision of information about things of a certain type (e.g. sustainability standard, company) that allows for comparability between those things.

Entity

- In this context, the subject of a sustainability benchmark. This can be a sustainability standard, company, NGO or other stakeholder.

5. Purpose of Benchmarking

Benchmarking is an important tool for a number of reasons:

- It provides the user with evidence to recognise, use or support a given sustainability initiative or entity in order to achieve defined sustainability objectives;
- It promotes consistency of performance between the benchmarked entities;
• It improves transparency about the operations and performance of the benchmarked entities;
• As a result, it helps to build awareness and uptake of these entities;
• At its best, it has the potential to improve the rigour and effectiveness of the benchmarked entities; and
• It creates a guidepost for stakeholders and other users to know what acceptable practice looks like.

While there may be good reasons to develop a new benchmarking programme, the default approach should be to first determine whether it is possible to reference existing benchmarking initiatives, either in whole or in part. New benchmarking initiatives should only be developed where an unmet need has been identified through a formal needs assessment.

Benchmarking is used by multiple actors for multiple purposes. Defining the purpose of the benchmark is important (see 8.1) and differences in organisational or policy objectives explain why a variety of benchmarks are developed. The table below illustrates some generic uses of benchmarks by different stakeholders.

<table>
<thead>
<tr>
<th></th>
<th>Governments</th>
<th>Companies</th>
<th>NGOs</th>
<th>Finance Institutions</th>
<th>Sustainability Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation mechanism to fulfil regulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainable procurement policies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency for users of sustainability initiatives or entities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk filter for investment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fulfil due diligence obligations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition of national standards in international markets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set expectations of benchmarked entities – drive consistency and improvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It is critical at the outset of any benchmarking exercise or programme to clearly define what is the purpose or goal of the programme and what the programme wants to achieve, as this will inform how the programme is structured.

It is important to note that benchmarking programmes do not always achieve positive results and may have unintended negative consequences. For example, where a benchmarking programme is implemented poorly it may result in recognition of poor performing entities and thus encourage a race to the bottom.
6. Typology

Benchmarks can apply to a wide range of entities. In the context of sustainability, the most common types of benchmarks evaluate an entity’s sustainability policies, practices or performance, or they evaluate the tools that those entities use to support their sustainability objectives (e.g. sustainability standards or similar tools).

Comparisons are precursors to benchmarks and consist of reviewing information about similar entities. Once a reference point is defined or a value judgement made about that information, then the comparison becomes a benchmarking exercise. That benchmark can be applied in different ways, the most common of which are for:

- **Recognition**: focused on qualifying entities that meet a threshold defined by the benchmark.
- **Ranking**: focused on ordering the performance of entities against the benchmark in relation to their peers
- **Improvement**: focused on gap analysis with an aspirational performance benchmark, to guide further development or strengthening of the subject

It is also possible to develop benchmarks that are a combination of these models, such as when a recognition benchmark also includes mechanisms for improvement.

**Benchmarking Typology**

![Diagram 1: Benchmarking Typology. For more information on the examples, please see Annex 1 (under development)](image)

**Comparison**

This first category includes programmes that aim to provide comparable information about the relative strengths and weaknesses of different entities. This comparison approach has value for informing users about different entities, enabling them to make choices about which ones to use or accept, based on their own interests or preferences. This approach is less normative than benchmarking approaches.
Recognition

This is the most common form of benchmarking exercise or programme since it enables the convenor to establish a common baseline for recognition. That recognition can have multiple applications and incentives attached to it, from governments recognising standards in their allocation of subsidies, procurement contracts or company due diligence obligations, to finance sector investment risk screens, to companies having clear information about which standards are acceptable for use in their supply chains.

Since recognition can serve different purposes, it follows that the nature of the baseline, or performance threshold, will vary significantly between benchmarking exercises. Benchmarking programmes will need to choose whether the performance threshold should be ‘broad and shallow’, covering a range of issues superficially, e.g. for determining some broad concept of market acceptability, or whether it should have ‘deep and narrow’ thresholds that can be very specific and focus on one priority area like modern slavery or child labour.

The implication for developing a comparison model is that it should aim to include the full range of issues and approaches covered by the benchmarked entities, so that users get a sense of the diversity of the entities’ performance and approaches. The challenge inherent in this is that comparison models run the risk that users equate ‘the best’ with those entities that tick the most boxes, whereas different criteria may be more appropriate for different contexts. Comparison models should educate users on the importance of determining which criteria or performance levels are relevant for their specific context.

Ranking

Rather than setting a performance threshold, this approach is focused on measuring the performance of entities against each other. The convenor still has to decide which criteria to use as the basis for the benchmark and how to weight or score those criteria, but rather than stating what are sufficient practices, the entities are evaluated and ranked in order of performance. This model is particularly prevalent in corporate benchmarks but can be applied equally to sustainability standards and other entities. There is some overlap with improvement models to the extent that ranking of entities creates an incentive for those entities to improve their performance.
Improvement

Less common but no less important are benchmarking programmes that aim to create drivers for improvement of the benchmarked entities. These benchmarks are often structured as aspirational performance bars that may go beyond current practices but that provide a roadmap for expectations about future performance. As practices improve, these benchmarks can be revised upwards to reflect future aspirations on sustainability performance and to instil an ethic of continual improvement. In practice, these have been applied especially by NGOs and by governments to incentivise sustainability standards and influence the definition of acceptable practices.

The implication for improvement model benchmarks is that they should be structured as aspirational criteria, sometimes going beyond good practice to look at best in class. They reflect the benchmarking programme’s views of what best practice should look like. These benchmarks are regularly reviewed and revised as best practice evolves, and generally require a higher degree of engagement with the entities being benchmarked.
7. Principles

Not all benchmarking exercises or programmes will look the same. Different programmes will have different goals or objectives that inform their structure. Regardless of individual approaches, a number of core principles are applicable across all benchmarking exercises and programmes and can be used as a point of reference for decisions about how to develop and implement the programme. These principles are derived from the ISEAL Credibility Principles. Another good resource is the World Benchmarking Alliance Guiding Principles (see p.124).

Transparency

Relevant information is made freely available in an accessible manner.

- In the context of benchmarking, this means that interested stakeholders have access to information about the mission of the benchmarking programme, the benchmark criteria, how the benchmark was set and how it is being implemented, and the results of any benchmarking and what those results mean.

Rigour

Benchmarking exercises and programmes are structured and implemented in ways that are sufficient to produce quality outcomes.

- This means that the benchmark is sufficiently detailed and clear, while the benchmarking process is robust\(^1\) and implemented consistently by individuals who are competent for their roles.

Stakeholder engagement

Interested stakeholders have appropriate opportunities to participate in and provide input to the process.

- This means that programmes have identified interested stakeholders and should determine whether and how they can provide input to the development of the benchmark or to the subsequent benchmarking of standards, and to make those opportunities\(^2\) available to them.

Impartiality

Benchmarking exercises and programmes identify and mitigate conflicts of interest throughout their operations.

- This means that the benchmarking programme is not engaged in activities that would compromise the integrity of the benchmarking results. Where one organisation benchmarks other organisations against itself, this principle would imply that the results only be used for internal purposes.

---

\(^1\) A robust benchmarking process is one which is consistent with these guidelines

\(^2\) Appropriate opportunities could include providing input to the setting of the benchmark, providing feedback on the evaluation of entities against the benchmark, or having a complaints process available to stakeholders.
Efficiency

Benchmarking exercises and programmes are structured as simply as possible and avoid redundancies.

- This means that a new benchmarking programme should only be developed if its objectives can’t be fulfilled by existing initiatives, that the programme is not unduly complex, that the number of benchmarking criteria are fit for purpose and not overly prescriptive, and that the criteria are aligned as much as possible with the most relevant existing benchmarking initiatives.

Improvement

Benchmarking exercises and programmes are structured to incentivise better practices in the entities that they cover.

- Ideally, this means benchmarking programmes stimulate a ‘race to the top’ rather than a drive towards the lowest common denominator, while recognizing that ‘better’ practices do not necessarily imply higher performance levels or more stringent requirements. Benchmarking programmes support mechanisms such as gap analyses that encourage and reward improvements in benchmarked entities, and revise and improve the benchmark itself, based on learnings from its implementation.

Accessibility

Benchmarking exercises and programmes avoid structures that create unnecessary barriers to participation and implementation.

- This means that the benchmark and accompanying procedures are appropriate, easy to understand, and broadly applicable, and that any associated fees or other requirements do not create significant burdens that would prevent or inhibit participation.
8. Benchmarking Development Process

The process to develop a benchmarking exercise or programme requires consideration of various design elements that will determine how the programme is structured. This section provides practical guidance for each of these elements, highlighting decisions that need to be made and the implications of different choices.

8.1 Determine the purpose of the benchmark and who it is for

8.1.1 Audience

The first step in thinking about a benchmarking programme is to identify who is the intended audience and what are their needs. Government agencies interested in setting sustainability guidelines for procurement may have significantly different needs and expectations from NGOs wanting to strengthen performance requirements in sustainability standards. Identifying primary and secondary audiences and their needs will influence how the benchmarking programme is structured, what benchmarking model is used, and the performance level at which the benchmark itself is set. A useful step is in understanding user needs is to speak with potential users directly to understand more deeply their expectations.

8.1.2 Purpose

Once a new initiative has identified its audiences and their needs, the next step is to define the goal or purpose of the benchmarking exercise or programme. Programmes need to have a clearly articulated purpose that reflects what they have learned about their users’ needs.

The purpose of any benchmarking programme should be made explicit and included in all communication about the programme. This will also help to inform the types of claims that benchmarked entities and users of the benchmarking programme can make about the results.

Convenors should also assess whether a benchmarking programme is the most effective strategy to achieve their purpose. Part of that assessment should include whether existing benchmarks already meet all or part of the articulated goal. A reference list of sustainability benchmarking programmes is included in Annex 1 (under development).

8.2 Determine who or what is being benchmarked

Within the scope of sustainability benchmarks, there can be a wide range of entities being benchmarked, from companies and corporate performance to the sustainability standards and related tools that companies use to convey their sustainability credentials. A decision about who or what is being benchmarked will greatly influence the setting of the benchmark content. The following diagram shows a series of choices that need to be made to arrive at a definition of the scope:
8.2.1  Type of entity

Once the sector and geography have been determined, the principle decision a benchmarking programme or exercise needs to make is whether the benchmark will focus on evaluating sustainability standards and initiatives or on corporate action towards sustainability, or some combination of these. This decision is driven by the defined purpose or goal of the programme, as the following examples highlight:

- To give clear guidance on how companies can operationalize the OECD’s Due Diligence Guidance for Responsible Business Conduct, the OECD is carrying out ‘alignment assessments’ or benchmarks for different sectors. In 2018 the OECD’s benchmarking programme on mining and minerals concluded its first alignment assessment of six related industry initiatives and sustainability standards, assigning them into one of three categories (not aligned, partially, or fully aligned) based on its benchmarking score. The benchmarking also resulted in a gap assessment that was used to strengthen alignment of these standards with the OECD Guidance.

- The World Benchmarking Alliance aims to make transparent the performance of the top 30 capitalised companies in each sector they focus on, by evaluating publicly available information about the companies against a benchmark derived from applicable SDG targets. The goal is to use peer performance as an incentive for companies to improve their sustainability practices and commitments and to give investors a transparent SDG performance ranking.

- The German government’s Green Button initiative is an unusual hybrid in which the benchmarking programme seeks to evaluate both a company’s performance and the sustainability credentials of its products, based on their compliance with responsible production standards. This approach can be achieved by combining what is essentially two benchmarking exercises.

8.2.2  Supply chain scope

The second decision to be made, appropriate primarily to benchmarks of sustainability standards, is what supply chain scope will be covered by the benchmark? Benchmarking programmes can focus on standards that
apply only at the production phase of the supply chain (e.g. deforestation legality) or can look at performance across all stages of the supply chain (e.g. labour standards).

Where benchmarks aim to recognise products that derive from standards-compliant production, the benchmark also needs to consider the chain of custody systems that are in place to ensure the integrity of the product through the supply chain.

8.2.3 Market presence

The third decision focuses on which entities to include in the benchmark. For a corporate benchmark, this means determining which types of companies to include. Is the benchmark focused only on companies of a certain size, sector, market presence, capitalisation, etc.? A similar exercise is necessary where the benchmark focuses on sustainability standards and similar tools. Decisions need to be made about the types of standards, sectors to which they apply, and scope.

8.3 Determine who will manage the benchmarking process

Benchmarking programmes are set up by many different types of institutions, from governments to companies and NGOs to the finance sector. In some cases, the institution that is seeking to use the results of the benchmarking programme is the one to develop and manage it. This is most often the case for companies seeking to understand which standards are most relevant for their supply chains. This model is also used where an entity seeks to compare their performance with their peers.

The advantage of this model is that the owner-operator has complete control over the content of the benchmarking, enabling them to tailor the process to their specific needs. However, the potential conflicts inherent in this model mean that it is not an appropriate basis for any public claims or communications. A variation on this model that is less conflicted is where the institution carries out the benchmark on its own behalf but has no formal affiliation to the entities being benchmarked. This would include examples such as the WWF Certification Assessment Tool and the World Benchmarking Alliance.

It is not always the case that the end-user of a benchmarking programme is the one to manage it. Most benchmarking programmes involve management of the benchmark being contracted out. This includes management by a related organisation or institution on behalf of the user, such as in the case of trade associations (SAI Platform or Global Sustainable Seafood Initiative (GSSI)), or management by a consultant, such as the contracting of Ecofys (now Navigant) to help the European Commission carry out their Renewable Energy Directive benchmarking.

It should be noted that even though a convening organisation may lack a direct interest in the outcomes, they still bring vested interests to the table. It is useful for any benchmarking organisation to declare its interests and any biases or intentional positions it brings to the benchmarking exercise. As one means to manage its interests, the benchmarking organisation should explicitly state the goal or purpose of the programme in all communications.
8.4 Determine the content of the benchmark

The benchmark itself is the reference against which the benchmarked entities are assessed. Given the diversity of potential entities to be benchmarked, it is not surprising that the content of the benchmark will differ significantly. For example, the types of criteria used to evaluate the sustainability performance of a retailer or manufacturer will look very different from the criteria used to evaluate the credibility of a sustainability standard.

This document does not prescribe where performance levels should be set for the content of a benchmark as these should be developed through a robust benchmark development process. Instead, we first describe the issues all benchmarking programmes will need to explore as they build out the content of their benchmark. Then, for sustainability standards in particular, we provide guidance for what elements of a standards system should be included in the benchmark and why these are important.

8.4.1 Considerations for developing benchmark content

8.4.1.1 Stakeholder engagement

Each benchmarking programme or exercise will need to decide how the content of its benchmark will be determined and to what extent the process of defining its content involves interested stakeholders. The value in considering stakeholder consultation comes both from the legitimacy and acceptance of the benchmarking process and from improving the relevance and feasibility of the benchmark itself.

The goal or purpose of the benchmarking programme is likely to inform the extent of stakeholder engagement. For example, where a government agency is setting a benchmark to assess fulfilment of a specific policy objective or where a company is determining which standards are consistent with its supply chain sourcing policies, there may be less scope for having stakeholders influence the content of the benchmark than where an industry association is seeking to establish minimum acceptable practices for sustainability standards operating in their sector.

Where stakeholder consultation or engagement is deemed appropriate, the following practices are useful to consider. Further information on good practices for stakeholder engagement are contained in the ISEAL Standard-Setting Code of Good Practice:

- Identify the types or categories of stakeholders who may be interested in the benchmark and the key stakeholders within each of those categories;
- Provide to stakeholders a clear and concise synopsis of the benchmark development process, how stakeholders can participate, and what happens to their input;
- Provide multiple different mechanisms through which stakeholders can share their opinions and feedback;
- Be transparent about how input is being taken into account and how decisions are made, potentially making available any comments received and responses;
- Test the applicability of the benchmark in the draft stage to ensure its feasibility and relevance;
• Make draft and final versions of the benchmark and benchmarking process freely available and easily accessible.

8.4.1.2 Technical experts

While stakeholders can provide meaningful input to help shape a benchmark, it is also important to have technical experts in lead roles preparing the benchmark language and taking stakeholder input into account. The choice that benchmarking programmes need to make is whether to contract consultants to carry out this work or to put in place a technical committee of experts. There are merits to either option and it is most important to define clearly what role these experts have in the decision-making process on the content of the benchmark. For example, do they play an advisory role, deferring to the convenor on decision-making, or are they the final arbiters of the benchmark’s content?

8.4.1.3 Conflicts of interest

For the legitimacy and acceptance of the benchmark, benchmarking programmes will need to manage for potential conflicts of interest. While the entities that will be subject to the benchmarking programme are keenly interested in its development, and can provide needed expertise, it is necessary to ensure adequate mechanisms are in place to avoid that these vested interests have undue influence over setting the benchmark.

8.4.1.4 Definitions

Part of the benchmark that is often overlooked is the definitions of specific terms. Much like any standard, evaluators who are assessing an entity’s alignment with a benchmark are going to have to interpret that benchmark and it is important they do so consistently. One way to support more consistency is to build definitions of key terms into the benchmark. Any terms that could potentially have multiple meanings or interpretations should be included in the list of definitions.

8.4.2 Core elements for benchmarking standards systems

Many benchmarking exercises or programmes for sustainability standards focus primarily on the content of the standard – the sustainability requirements that need to be met by certifying enterprises. This is clearly an important component of benchmarking, but not sufficient. Two standards that look identical on paper can support very different sustainability outcomes depending on how they are implemented in practice. It is, therefore, necessary to look at both the content of the standard and the system that supports its uptake. This guidance does not propose content requirements as these are specific to the context of the benchmark.

The two core elements that should form part of any benchmark of sustainability standards are:

• The standard - sustainability performance levels delineated in the standard. This may include environmental and/or social sustainability, covering issues such as environmental management, biodiversity, natural resources, social well-being, labour rights, and economic livelihoods. In addressing these issues, standards can focus on defining acceptable practices, processes that should be followed, or desired outcomes; and

• Assurance - the means of assessing a company’s compliance with the requirements in the standard. This is critical because assessment models vary significantly so that even when two standards look identical on paper, verification practices may lead to different compliance results. Assurance
elements most commonly incorporated include the certification or verification procedures, competencies required of auditors or evaluators, and accreditation or oversight of the assurance providers. The level of independence of the assurance process is also a consideration, distinguishing between self-assessments (first party), interested parties (second party), and independent entities (third party). Often, assurance models are constituted by a combination of these approaches. Examples of what credible assurance looks like are contained in the ISEAL Assurance Code of Good Practice.

It is possible that benchmarks can focus just on evaluating the content of sustainability standards and not the assurance mechanisms, but this severely limits the types of conclusions that can be drawn from the evaluations and is not recommended. For example, benchmarks focused only on standards content would preclude users from making any claims of comparability of the systems being benchmarked. It also prevents the benchmark being used as a way to recognize and incentivize companies or entities that are compliant with the benchmarked standards. The one instance where a limited scope for the benchmark may be warranted is in improvement model benchmarks that are seeking to improve standards’ requirements around specific sustainability issues but, even here, insight about how the standard is assessed in practice would be valuable.

Additional elements that are useful to incorporate into a benchmark of sustainability standards include:

- **Scheme governance** – how the standards system is structured and implemented. The legitimacy of a sustainability standard derives in part from its operating structure. This includes the extent to which stakeholders participate in the development of the standard and the governance of the scheme. It also encompasses the existence of checks and balances like the inclusion of robust grievance or dispute resolution mechanisms, and the extent to which the scheme makes information about its procedures and assessment results transparent and accessible.

- **Traceability (chain-of-custody)** – the extent to which products from a certified production unit or facility can be traced through the supply chain to the final buyer. There are a number of different traceability models of varying rigour (e.g. from identity preservation to book and claim) and each is appropriate for different purposes and enables different claims to be made about the end products. Benchmarks may stipulate the type of traceability required or could focus on increasing transparency around the variation in traceability solutions.

- **Sustainability claims** – the communication of performance by a certified enterprise. While not often included in the scope of a benchmark, there is a strong case to be made for the critical importance of assessing the appropriateness of the sustainability claims allowed within the standards scheme and by the participating enterprises. The legitimacy of a sustainability standard is based in large part on a meaningful correlation between the level of ambition of the sustainability practices or outcomes required and the commensurate sustainability claims allowed.

For an introductory overview of the issues and criteria that are useful to include in a benchmarking of standards systems, please see Annex 3 and the ISEAL Introduction to Comparing and Benchmarking Sustainability Standards Systems.
8.4.3 Alignment with international norms and guides

In determining which criteria to include in a benchmark, it is often the case that the benchmarking programme will be building off existing criteria, whether these are in the form of policy stipulations or company commitments. It is useful to consider referencing external criteria and frameworks when developing a benchmark, as this improves both the quality of the benchmark and the potential consistency between benchmarks.

For the sustainability performance criteria within a benchmark, be it for standards or companies, it is useful to refer to the products of inter-governmental initiatives as these have inherent legitimacy and global relevance. Among the most relevant and recognised frameworks to consider are the following:

- UN Sustainable Development Goals (SDGs)
- UN Guiding Principles on Business and Human Rights
- OECD Due Diligence Guidance for Responsible Business Conduct (general and sector specific)
- FAO Code of Conduct for Responsible Fisheries, and Guidance for Responsible Agriculture Investing
- ILO Core Labour Conventions and Fundamental Principles and Rights at Work

While these international frameworks are useful to structure and delineate benchmarking content, they still need to be transposed into comparison or benchmarking criteria that are clearly measurable and relevant to the scope and purpose of the benchmark.

For sustainability standards specifically, there are additional common reference points for good practice in how these systems are implemented, including:

- ISO Conformity Assessment standards (17000 series of standards) – for core certification and accreditation functions
- ISEAL Codes of Good Practice – for standard-setting and governance, sustainability assurance, and measuring impacts of sustainability standards.

8.4.4 Accommodating diverse approaches in a benchmark

One of the most significant challenges inherent in benchmarking is how to recognise a diversity of models or approaches in the benchmark, while at the same time being prescriptive enough to ensure its rigour. The implication is that initiatives can only be benchmarked on the characteristics that they share in common. For example, let’s use the analogy of a fruit bowl:

- If the fruit bowl contains only apples then there are a wide range of characteristics we can use to compare and benchmark the apples, including type of fruit, size, colour, firmness, taste, etc.
- If the fruit bowl contains apples and oranges then the number of characteristics decreases. Does it make sense to benchmark the size or colour of apples versus oranges? It may be that we can only assess whether all the items in the fruit bowl are actually fruit.
The point is that benchmarks need to compare like with like or find the common characteristics between different entities. The greater the consistency of entities being benchmarked, the more detail can be included in the benchmark and the stronger the claims that can be made.

Where a benchmarking programme includes entities with different characteristics, then the exercise is to identify all those characteristics that they have in common as a starting point for determining what is relevant to include in the benchmark. Not all common elements need to be included in the benchmark if they don’t support achievement of the programme’s purpose.

### 8.4.4.1 Variations in sustainability standards and related tools

In the case of sustainability standards, there is increasing variation in how they and similar tools are structured and how they seek to incentivise sustainability improvements. This presents a fundamental challenge for benchmarking since these tools are, by nature, establishing common reference points.

The way a standard is structured will influence how the benchmarking performance criteria are set. Standards themselves can take many forms, the most common of which are:

- those that delineate desired practices;
- those that define processes or systems that should be in place;
- those focused on desired outcomes that should be achieved; or
- a combination of the above.

Similarly, variations in the types of assurance models that are used to assess compliance with the standard will influence the choice of benchmark criteria. Among the most significant variations are differences in how compliance is measured or scored. For example, improvement model standards can recognise enterprises as compliant once they have started on the improvement journey, compared to standards that require enterprises to achieve a high performance bar before they are considered compliant.

A standards system can set the compliance bar at any level in relation to the content of the standard, making this as important a consideration for consideration in the benchmark as the content requirements themselves.

An additional challenge is that sustainability standards are now experimenting with different strategies for incentivising uptake of more sustainable practices. For example, they are delivering capacity building or are convening stakeholders for collaborative action. Benchmarks of sustainability standards don’t often recognise or compare these additional strategies and yet the strategies are becoming increasingly important to the understanding of how sustainability standards deliver systemic impacts.

Finally, it is also worth considering the variety of sustainability tools that are emerging that play similar roles to standards and certification, employing slightly different strategies to communicate sustainability performance of enterprises, such as geospatial mapping of deforestation or communication of improvements through performance dashboards.

If a benchmarking exercise or programme chooses to incorporate sustainability standards and tools that employ quite varied strategies for uptake, then there are two approaches the programme can take:
The first is to focus the benchmark primarily on the content of the standard or performance requirements, since all initiatives incorporate these to some extent. Being less prescriptive on criteria related to the implementation strategies will mean that a broader suite of tools can be assessed. However, as noted earlier, initiatives with similar standards can look very different when implemented in practice. This first approach is likely a better scenario for use with comparison models, since the goal would be to highlight the similarities and differences between initiatives, rather than to qualify one initiative or the other.

The second approach is to consider whether there are benchmark criteria specific to each general type of uptake strategy. Rather than have one set of criteria to evaluate all standards and similar tools, it may make sense to have multiple sets of criteria that are applied to different initiatives as appropriate. For example, those standards or tools that employ certification as a strategy would be evaluated against the certification benchmarking criteria; those that deliver training would be evaluated against training-related criteria, etc. Initiatives could be evaluated for more than one set of criteria if they employ multiple strategies.

It should be noted that capturing multiple strategies in benchmarking programmes is an emerging field, and little pre-existing experience is available to define what good practice looks like. Therefore, some amount of trial and error should be expected.

8.4.4.2 Capturing improvement models

There is a growing trend in sustainability standards and similar tools that causes a challenge for benchmarking: the shift from a pass/fail certification approach to a continual improvement model. This trend recognises the value and necessity of engaging enterprises at all levels of sustainability performance and creating the levers necessary to incentivise their improved performance over time. The challenge for benchmarking is that while the standard becomes the aspirational goal for participating enterprises to meet, each of those enterprises is at a different stage in their improvement journey. It is not possible to say that the programme delivers a certain performance level across all participating entities. This is similar to the challenge of standards that have different scoring approaches, where e.g. a standard may only require that 80% of the criteria are met for the enterprise to be certified, without specifying which 80%.

Traditionally, benchmarking programmes have penalised improvement standards by evaluating only those minimum criteria that they are certain are being met by all participating enterprises (the core criteria included in most improvement standards). This is a logical approach if the goal is to convey what we know for certain about performance levels. However, a more nuanced approach is to carry out two assessments, of both the minimum performance level and the level required to be met over time, and then include whether and how quickly that upper performance level must be reached. For example, many improvement standards require that participating enterprises make regular progress each year, moving to full compliance with the standard within a set period. That would be considered more rigorous than having a baseline set of practices but not requiring any improvement towards aspirational goals. At minimum, benchmarking programmes should distinguish which standards include improvement approaches so that these can be compared with other like programmes.
8.5 Determine the evaluation structure of the benchmark

An important component of the benchmarking exercise or programme is a decision on the evaluation structure. This decision is also linked to the purpose of the programme (see 8.1.2). The simplest evaluation structure for a benchmark is that all criteria are equally weighted and mandatory. However, there are numerous models or derivations of this approach to suit different circumstances:

- **Mandatory vs aspirational** – a distinction between mandatory or core criteria and those that are aspirational or advanced. In a recognition model benchmark, a clear delineation needs to be made between those criteria that are required to be met to qualify and those that enable differentiation above and beyond the core requirements.

- **Scoring** – even where all criteria are mandatory, there is an option to implement a scoring protocol for qualifying the benchmarked entities. There are many variations in potential scoring models, the simplest of which would require that x% of the criteria are met in order to qualify. Alternatively, the scoring model could require that all core requirements are met as well as a percentage of the aspirational requirements.

- **Ranking** – this is a variation on the scoring model where the entities are not qualified against a set score but are instead ranked against each other. This is common in corporate benchmarking and also creates an incentive for an initiative to improve its performance over time relative to others.

- **Progress models** – as with improvement standards, it is possible for benchmarks to incentivize progress over time. From the perspective of the evaluation structure, this means differentiating between mandatory and aspirational criteria and increasing the number of requirements that are mandatory over time. Alternatively, it could mean carrying out regular revisions of the benchmark to increase the mandatory criteria as good practice evolves.

As with other aspects of the benchmarking process, no matter which evaluation structure is applied to the benchmark, a critical element is to be transparent about what this model is and how it works, with good practice being to publish the benchmarking methodology in full.

8.6 Determine the benchmarking implementation process

8.6.1 *Steps in the benchmarking implementation process*

The second core component of a benchmarking exercise or programme is determining how the programme will be implemented. The steps in the benchmarking process are relatively straightforward and tend to be similar across most benchmarking initiatives:
**Application** – the first step in the process is for the benchmarking programme to decide whether to intentionally choose which entities to evaluate or create an application process by which any qualifying entity can apply for evaluation.

Recommendations:

- The process could begin with a self-evaluation, giving qualifying entities the opportunity to assess how aligned they are with the benchmark before formally applying to be evaluated.
- Where the entities are chosen by the benchmarking programme, it is important to reach out to them at an early stage to establish good lines of communication and build trust in the process. In the best cases, non-voluntary benchmarks can be an incentive for benchmarked entities to share performance information beyond what is publicly available.
- The desired outcome from the application stage is that the benchmarking programme has all the information about the entity necessary to conduct an evaluation. A checklist can be developed that lists the types of documents or information that should be made available.
- If there is a fee payable by the entity to participate in the benchmarking, it is appropriate to collect it at this stage in the process.

**Desktop review** – the main step in the benchmarking process is the evaluation of how the entity is aligned with the benchmark. This consists most commonly of a desk review of detailed documentation about the entity’s procedures and practices.

Recommendations:

- A consistent evaluation procedure (the benchmarking protocol) and reporting format should be developed to guide the review process. This procedure should include all the steps in the evaluation and should be made publicly available.
- Most importantly, benchmarking programmes will need to decide who will carry out the reviews and ensure that these individuals or organisations are competent for the task. More recommendations on this element are included under 8.8 Further Considerations.
- Where the entities are chosen by the benchmarking programme, evaluations could be based initially on publicly available information, then supplemented by information shared by the entity. This additional input will be helpful in gaining a more complete understanding of how the entity operates.

**Additional data collection** – this is an optional step, included in some benchmarking processes to give greater insight about how the entity operates in practice. The general premise is that multiple sources of information will provide a better picture of potential performance. Among the means of additional data collection are office visits, often used in the case of sustainability standards; interviews with relevant stakeholders, such as NGOs or community organisations impacted by a company’s operations; and stakeholder comment platforms. These additional activities will help to identify and bring to light entities that may have good procedures in place but little effort to implement those procedures.
Recommendations:

○ This step is optional because of the additional cost involved. In general, the costs of benchmarking processes are a major constraint and inhibit accessibility, so an office visit, in particular, is only advised where the benchmarking programme is covering the costs or where a deeper level of trust-building is required, such as where benchmarking is a step towards mutual recognition between entities.

○ Witness audits are an additional level of scrutiny that can be included for evaluation of sustainability standards, in addition to or instead of an office visit. A witness audit involves the evaluators witnessing an assessment by the sustainability standard’s assurance programme, to understand better how compliance with the standard is evaluated in practice.

○ Much like the desktop review, an office visit should be conducted according to consistent operating procedures that are made publicly available.

• **Benchmarking committee** – benchmarking committees can serve as a useful complement to the evaluators, by providing a second set of eyes for the evaluation and bringing a level of consistency to interpretation of the criteria by different evaluators. The first responsibility of a committee would be to review the work of the evaluators and ensure the evaluations are ready for public scrutiny.

Recommendations:

○ To manage costs, the committee can be small in number (e.g. 2 or 3 people). The most important factors are: consistency in who participates and the qualifications of the committee members, their independence, and their knowledge about the benchmark and the sector.

○ Where a committee is not put in place, the evaluators could also function as a peer group to review each other’s work, aiming for a similar level of consistency in interpretation.

• **Public consultation** – an optional public consultation consists of making the draft evaluation publicly available for a period of time (e.g. 30 days) so that interested stakeholders can provide feedback, both on the evaluation and on insights they have about the benchmarked entity.

Recommendations:

○ While this step is also optional, it is highly recommended in most cases as a means to gather additional insight about the entity, providing a useful complement to the desk review, particularly where an office visit or witness audit is not included.

○ If the benchmarking programme has built a strong and open relationship with the benchmarked entity, then the entity could help by raising awareness with interested stakeholders about the consultation.

○ Consultation input should be considered as one source of information to be weighed alongside other information, like the evaluation results. It is important to assess the potential for bias or conflicts of interest in the feedback received through consultation.
The benchmarking programme can decide whether to make input received publicly available, along with a summary of how that input was considered. Good practice favours more transparency than less.

- **Evaluation decision** – in most benchmarking exercises or programmes, and recognition model benchmarks in particular, the process will result in a decision about the extent of alignment with the benchmark.

  Recommendations:
  
  - The benchmarking programme will need to determine who makes the decision. Is it the evaluator, a committee, or the benchmarking programme staff? As with the evaluation, the key considerations are consistency and competence in who makes the decision, and transparency about the decision-making process and results.
  
  - Decisions do not always need to be about alignment. In comparison models, the decision could be about finalizing the assessment so that information about the subject can be made publicly available.
  
  - In coming to a decision, different sources of information will be used, some of which may conflict. It will be important to consider the source, robustness and integrity of each type of data or information to assess how much weight to assign to it in the decision-making process.
  
  - Notifying the benchmarked entity in advance of making any decisions public is good practice, providing the opportunity to discuss any findings and clarify any misconceptions.
  
  - It is also useful to have in place a dispute resolution mechanism, in cases where the perspectives of the benchmarking programme and the entity can’t be reconciled.

### 8.6.2 Monitoring of continued alignment

The operations and practices of both the benchmarked entity and the benchmarking programme itself can evolve over time, requiring a regular re-evaluation of alignment between the two.

  Recommendations:
  
  - The benchmarking process can be arduous and expensive so it is important to balance the need for up-to-date evaluations with the extent of effort required. It makes the most sense to re-evaluate entities after the benchmark has been updated and for that to happen on a regular basis but no more frequently than every three years.
  
  - A re-evaluation should be more streamlined than the original evaluation and structured so that the original evaluation acts as a starting point on top of which the benchmarked entity can provide any information about changes in its procedures and practices and any information relevant to new or revised requirements within the benchmark.

### 8.7 Determine how the results will be communicated

The final area to consider in the development of a benchmarking programme relates to the claims and other communications about the programme and about the results of evaluations. This is another factor that is
critical to the credibility and legitimacy of the benchmarking programme. Claims need to be grounded in, and consistent with the actual results and should not misrepresent the conclusions that can be drawn from the evaluation.

It is important to consider at the outset of the development process the types of claims that the benchmarking programme and the benchmarked entities can make since this will inform how the benchmark and benchmarking process are structured. For example, if the desired claim is about the entity meeting a benchmark, then the benchmarking programme would likely need to construct a recognition model benchmark. Conversely, if the benchmarking programme wants to avoid compliance claims then it may be more appropriate for them to consider setting up a ranking model benchmark or a comparison model.

All benchmarking programmes should have a defined claims policy that is publicly available and that covers both claims that the programme can make and claims made by the entities regarding their engagement in the benchmarking. It should also cover management protocols and ramifications in the case of misuse or mis-communication of the benchmarking results. The policy should make it clear who can say what about benchmarking results at what point in time. It should also define the lifespan of applicable claims, recognising that benchmarking results and the claims that depend on them can become out of date after a period of time. Additional guidance is available through ISEAL’s Sustainability Claims Good Practice Guide.

Communication of benchmarking results should strive for simplicity. This is about making it easy for stakeholders to understand how the benchmarking process came to its conclusions and what those conclusions mean. For example, if a benchmark includes evaluation of both mandatory and aspirational criteria, the benchmarking programme should communicate clearly what this means and how benchmarked entities can be differentiated. The proposed language should be tangible and concrete, clearly describing what the results mean and ideally tying them to the expressed goals of the benchmarking programme. Additional guidance on core principles for sustainability claims can be found in the UNEP Guidelines for Providing Product Sustainability Information.

8.8 Further considerations

8.8.1 Competence

The legitimacy and value of the benchmarking process depends on the competence of all those involved to implement the process professionally and consistently. This is particularly the case for evaluators and decision-makers.

Recommendations:

- Define a set of skills and competencies required for each position in the benchmarking process and ensure that personnel or contractors meet those competencies.
- It is usually more effective to have a small group of well-trained evaluators than a large group, because it is easier both to maintain consistency of interpretation between a small
group and to support evaluators maintaining a deep understanding of the benchmark and its intent over time.

○ Where evaluators or decision-makers need to make personal judgements, these should be noted and ideally incorporated into a systemic approach that seeks to minimize the subjectivity. However, realistically, it is also equally important for the benchmarking programme to be comfortable with some level of ambiguity.

Evaluation results will almost always include some level of subjectivity, given the potential for differing interpretations of the benchmark. It is important both to ensure there is clarity of intent behind each element in the benchmark to inform how the evaluators should interpret it, and to document how different criteria are interpreted to build a body of knowledge for subsequent evaluations.

8.8.2 Cost and complexity

As noted in a few places throughout this guidance, benchmarking processes can be quite burdensome for the entities being evaluated as well as for the benchmarking programme. Benchmarking programmes need to find a balance of costs and complexity that achieves meaningful results while still being manageable.

› Recommendations:

○ High profile, international standards and multinational corporations, in particular, are the subject of numerous benchmarking processes. The most effective way to reduce cost and complexity is for a new benchmarking initiative to use or adapt existing benchmarks rather than create a new one.

○ Fostering collaboration between benchmarking initiatives is also valuable to increase consistency on what is being evaluated. This will help to reduce already existing duplication. Not all benchmarking requirements can be harmonized since different benchmarking programmes have different goals and objectives. However, catalysing conversations fosters a better understanding and alignment between benchmarks.

○ It is good practice for benchmarking programmes to cover the costs of carrying out the evaluations. This should be budgeted for in the development of the programme, as should ongoing costs to maintain and update the benchmark and conduct re-evaluations. An exception to this is where successful alignment with a benchmark provides an entity with significant new market access.

8.8.3 Improving the benchmark and process

Through implementation of the benchmarking process, it is likely that the benchmarking programme will gain insights about what works well and how to improve both the benchmark and the process. A benchmarking programme is dynamic and should be revised regularly to reflect these insights.

› Recommendations:
• It is easier to capture insights and learning if a good data management system is in place. Some forethought should be given to how the results of evaluations and other data will be stored and managed for easy analysis. Being able to analyse where entities are most aligned or not with the benchmark is an important foundation for a data-driven revision. Similarly, capturing data like time and cost allocations at different stages of the benchmarking process may highlight areas where increased efficiencies are possible.

• As noted above, the benchmark should be updated at regular intervals but not so frequently that it creates additional cost burdens. Every three to five years is recommended.

8.8.4 Government use of benchmarks

Technical, political and/or economic obstacles might affect how public bodies implement certain benchmarking practices and how they integrate benchmarking results into policies that support sustainability.

Recommendations:

• Where the results of a benchmark are tied to incentives for companies, such as access to subsidies, public procurement contracts, or exemption from government inspections, there is more pressure on the benchmarking initiative to make its process and evaluation decisions fully transparent in order to mitigate for potential conflicts of interest.

• In certain legal contexts, such as EU public procurement, governments need to accept products or services that are deemed ‘equivalent’. For benchmarking, this means that a government body will need to consider any company’s claim to meet their requirements and cannot choose to recognise only one or a few specific standards.

• Government policies that incorporate sustainability benchmarking might need to consider the supply of products that would meet the benchmark. If a policy affects imports, the government could be accused by exporting countries of setting up illegitimate barriers to trade. Some experience suggests that overly ambitious demand-side policies (e.g. banning all palm oil imports unless they are certified against acceptable standards) are politically unfeasible. One option to address this is to use broader benchmarking and recognition criteria, such as including improvement models, and then ratcheting these up over time so that export markets have time to adjust.
Annex 1: Summary of Sustainability Benchmarking Programmes

(to be completed)

The following table provides a quick reference overview of different existing sustainability benchmarking programmes. It is recommended good practice to determine whether an existing benchmarking programme can meet all or part of a new initiative’s expressed needs, rather than developing a new benchmarking programme. This short summary is complemented by a series of Fact Sheets about these benchmarking programmes that will be produced by the UN International Trade Centre (ITC) in the context of their Sustainability Map initiative.

<table>
<thead>
<tr>
<th>Benchmarking programme</th>
<th>Audience</th>
<th>Purpose</th>
<th>Who or what is benchmarked?</th>
<th>Scope of benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMZ / GIZ SSCT Kompass Nachhaltigkeit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMZ Green Button</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGF Sustainable Supply Chain Initiative (SSCI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Sustainable Seafood Initiative (GSSI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDH Floriculture Sustainability Initiative (FSI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITC Sustainability Map</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD Benchmark on Guidelines for multinational enterprises</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>World Benchmarking Alliance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWF Certification Assessment Tool (CAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others??</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annex 2: Benchmarking Good Practice Checklist

This checklist distils key steps in the guidance for setting up and implementing a sustainability benchmarking programme. It is intended to be used as a quick reference both for those organisations and initiatives that are setting up new programmes and for stakeholders who want to assess the credibility of existing programmes and to hold them to account. The checklist should be used in conjunction with the full guidance.

Develop the framework

1. **Audience**: identify who is the intended audience and what are their needs; speak with potential users to understand more deeply their expectations (8.1.1)

2. **Purpose**: define the goal or purpose of the benchmarking programme and ensure this is explicit and included in all communication about the benchmarking programme (8.1.2)

3. **Strategy**: assess whether a benchmarking programme is the most effective strategy to achieve their purpose (8.1.2)

4. **Needs assessment**: assess whether existing benchmarks already meet all or part of the articulated goal (use Annex 1 as a starting point) (8.1.2)

5. **Scope**: make a decision about who or what is being benchmarked, including the sector or commodity, geography, type of entity, supply chain scope, and market presence (8.2)

6. **Openness**: decide whether to target specific companies or initiatives with the benchmark and which ones, or to allow any qualifying entity to be evaluated (8.2.3)

7. **Management**: determine who will manage the benchmarking programme (8.3)

8. **Claims**: develop a publicly available claims policy that ensures claims are grounded in, and consistent with the actual results and do not misrepresent the conclusions that can be drawn from the evaluation (8.7)

Determine the benchmark

9. **Process**: set the process for determining the content of the benchmark (8.4)

10. **Stakeholders**: determine whether and how to engage stakeholders in the content development and, where appropriate, follow good practices in the ISEAL Standard-Setting Code for how to engage stakeholders (8.4.1.1)

11. **Technical experts**: define the role for technical experts in content development, including their role in decision-making (8.4.1.2)
12. **Commonalities**: identify all the characteristics that entities that fall within the scope of the programme have in common, as a starting point for determining what is relevant to include in the benchmark (8.4.4)

13. **References**: use international reference documents to inform content and encourage consistency (8.4.3)

14. **Definitions**: include definitions of key terms in the benchmark to support consistent interpretation (8.4.1.4)

15. **Content**: for sustainability standards, include in the benchmark at least standards content criteria and criteria to assess the assurance mechanisms. Also strongly consider criteria on governance, traceability and sustainability claims (8.4.2 and Annex 3)

16. **Alternative models**: for sustainability standards and related tools, consider how to accommodate different standards models, including those with different scoring models, different assurance models, and different strategies for incentivising uptake of more sustainable practices (8.4.1.1, 8.4.4.2)

17. **Evaluation**: determine the evaluation structure of the benchmark (8.5)

**Develop the benchmarking process**

18. **Effectiveness**: find a balance in the benchmarking process that achieves credible results in an accessible and cost-effective way (8.8.2)

19. **Application**: determine the application process where benchmarking programmes are open to qualifying entities (8.6.1)

20. **Desk review**: carry out a review of detailed documentation about the entity’s procedures and practices (8.6.1)

21. **Additional data collection**: consider whether and how to gather additional information, such as an office visit or witness audit, to inform potential performance (8.6.1)

22. **Benchmarking committee**: determine whether to put in place a benchmarking committee or some other mechanism (e.g. evaluator peer review) to support consistency of interpretation (8.6.1)

23. **Public consultation**: consider a public consultation on draft evaluations and put in place the steps to do so where relevant (8.6.1)

24. **Decision-making**: determine how decisions on benchmarked entities will be made (8.6.1)

25. **Dispute resolution**: put in place a dispute resolution mechanism (8.6.1)

26. **Alignment**: establish a process for monitoring continued alignment between the benchmark and the benchmarked entity over time (8.6.2)

27. **Competence**: ensure that evaluators, decision-makers and others involved in the benchmarking process are competent for their work (8.8.1)
Additional considerations

28. **Transparency**: make information about the benchmarking programme, including how it works, governance, policies, decision-making and results publicly available and accessible (various clauses)

29. **Impartiality**: manage for potential conflicts of interest in setting of the benchmark and implementation of the benchmarking programme (8.3, 8.4.1.3, 8.6.1, 8.8.4)

30. **Improvement**: capture insights and learning from implementation of the benchmarking programme to inform its regular revision and improvement (8.8.3)
Annex 3: Benchmark Criteria for Evaluating Standards Systems

This list of criteria represents a summary of core operating practices that a sustainability standards system should have in place in order to be considered credible. These practices derive from ISEAL Codes of Good Practice and credibility tools, which represent a broadly shared understanding of what good practice looks like. With this list, ISEAL offers a snapshot for stakeholders to better understand what to look for when evaluating the likely effectiveness of sustainability standards. This list can also be used as a basis or starting point for defining the systems implementation criteria to be included in a benchmark.

Scheme Management

1. The scheme owner has a sustainability-oriented mission or vision
2. The scheme owner has defined and makes publicly available its desired long-term sustainability impacts and strategy for achieving those impacts
3. Stakeholders have an opportunity to provide input on the intended sustainability impacts and possible unintended effects of the standards system
4. On a regular basis, the scheme owner monitors and evaluates progress towards its sustainability impacts and accurately and publicly communicates the results
5. The scheme owner applies adaptive management by using the learning from monitoring and evaluation to improve its standard and supporting strategies
6. The scheme owner carries out internal or external audits of its management system and operations at least annually and incorporates the findings
7. The scheme owner makes information on the governance structure and income sources or financing structure of the scheme publicly available
8. Stakeholders have the possibility to participate in or provide formal input on the governance of the scheme

Standard-Setting

9. Information is made publicly available on standards development and revision processes and on decision-making
10. Consultations on the development or revision of the standard are open to all stakeholders
11. Input received during consultations is documented and there is a public report back on how issues raised are addressed
12. Decision-making on the content of the standard includes a balance of stakeholders and aims for consensus
13. The standard and consultation drafts are made freely and publicly available
14. Criteria in the standard contribute to addressing the key sustainability hotspots for the scope of the standard

15. The structure of the standard or accompanying guidance ensures consistent interpretation (e.g. auditable indicators)

16. There are provisions or mechanisms to ensure that the standard is locally applicable in the regions where it is applied

17. The standard is reviewed and revised on a regular basis (not exceeding five years)

**Assurance**

18. The overall assurance methodology and structure for the scheme are publicly available

19. Assurance bodies are required to implement a management system that supports consistency, competence and impartiality (e.g. ISO 17065, 17021 or equivalent)

20. Full audits of at least a sample of clients are carried out regularly (from every year to every 5 years depending on sector)

21. Full audits include office visits and on-site assessments of at least a sample of operations

22. Stakeholders have an opportunity to provide input to the audit

23. There is a publicly available methodology for how compliance with a standard is determined (e.g. the scoring methodology)

24. Decision-making on compliance is impartial

25. There are procedures for how clients are required to address non-compliances, including when a certificate is suspended or revoked

26. There is a publicly accessible complaints and appeals process for certification decisions

27. Summaries of certification assessment reports are made publicly available

28. The certificate or license defines the scope of certification and duration of validity

29. A list of all certified enterprises is made publicly available

30. The scheme owner carries out regular reviews of its assurance programme and notifies assurance bodies and clients of any changes in requirements

**Group Certification (where applicable)**

31. Groups are required to operate an internal management system that includes procedures for inducting, evaluating and removing group members

32. There is a representative sampling methodology for assessing group members during the external audit, and defined repercussions when a sampled member is found to be non-compliant

**Personnel Competence**

33. Specific qualifications and competencies are defined for auditors and assurance body personnel
34. Auditors and assurance body personnel are required to have an in-depth understanding of the standard and its interpretation

35. New auditors have a probationary period during which their competence in an audit is assessed or supervised

36. Auditors and assurance body personnel are required to participate in regular training and professional development

37. The competence of auditors and assurance body personnel is demonstrated through regular evaluation

38. There are repercussions such as probation or suspension for misconduct or poor performance of auditors or assurance body personnel

Oversight

39. There is an oversight mechanism that is independent of the assurance bodies being assessed

40. There are documented procedures for oversight and a management system that ensures consistent and competent application of these procedures

41. Individuals involved in oversight are competent to evaluate assurance bodies and possess knowledge of the standard and its intent

42. Oversight includes a review of the performance of assurance providers and auditors in the field

Chain of Custody (where applicable)

43. Chain of custody verification is required if the scheme results in a communication of product origin from certified production

44. All enterprises that physically take products into storage are assessed (except where handling tamper-proof packaged products)

45. Enough information is documented in the chain of custody assessment to enable tracing of the product and to avoid fraud in the supply chain

Claims and Labels

46. There are publicly available requirements for the use of claims and labels, including minimum levels of certified product content required for use of claims

47. A legal agreement is required for the use of claims and labels by enterprises in the supply chain

48. The types of claims allowed are appropriate considering the chain of custody models being applied

49. Allowable claims and labels contain enough information that their validity can be checked

50. The scheme employs surveillance strategies to monitor and rectify misuse of claims and labels