PARTNERSHIP FOR MARKET READINESS

TECHNICAL NOTE 14 | APRIL 2017

A Guide to Greenhouse Gas Benchmarking for Climate Policy Instruments

WORLD BANK GROUP

A Guide to Greenhouse Gas Benchmarking for Climate Policy Instruments

Cover photo: courtesy of istock.com

Cover design: Gregory Wlosinski

Acknowledgements

This Guide was prepared for the Partnership for Market Readiness, jointly by a team of experts from Ricardo Energy and Environment and NewClimate Institute for Climate Policy and Global Sustainability. Carolina de Oliveira and Mark Johnson led the team from Ricardo Energy and Environment, and Carsten Warnecke and Ritika Tewari led the team from NewClimate Institute for Climate Policy and Global Sustainability.

Harikumar Gadde and Pauline Maree Kennedy (World Bank), provided substantive inputs and managed the project.

We sincerely thank representatives from climate policy jurisdictions who shared their practical insights and knowledge related to designing and implementation of benchmarks for climate policy instruments through a survey, interviews, and review of the Guide. These include Clare Lonergan and Daniel Besley (Australia), Mary Jane Coombs and Mihoyo Fuji (California), Sebastian Carranza (Columbia), Alexandra Manole (European Commission), Alexandra Zirkel, Frank Gagelmann, Dr. Markus Kollar and Christoph Kühleis (German Emissions Trading Authority), S. Vikash Ranjan (India), Akihisa Kuriyama (Japan JCM), Zhibek Issakyzy (Kazakhstan), Víctor Escalona Gómez (Mexico), Ted Jamieson (New Zealand), Memory Machingambi and Sharlin Hemraj (South Africa), Sachiko Nakamura (Tokyo), Torsten Greis and Lea Bigot (Tunisia), Nicolas Muller (UNFCCC), and Vintura Silva (UNFCCC).

.

Contents

List of Boxes

List of Figures

List of Tables

List of acronyms

BAT	Best Available Technology
BAU	Business as Usual
BOF	Basic Oxygen Furnace
BREF	Best Available Technical Reference
CA	California
CCA	Climate Change Agreement (United Kingdom)
CDM	Clean Development Mechanism
CT	Carbon Tax
EAF	Electric Arc Furnace
EETS	Energy Efficiency Trading System
ETS	Emissions Trading System
EU	European Union
GHG	Greenhouse Gas
JCM	Joint Crediting Mechanism (Japan)
MMR	Mandatory Reporting Regulation
MRV	Monitoring, Reporting and Verification
NAICS	North American Industry Classification System
NAMA	Nationally Appropriate Mitigation Actions
NDC	Nationally Determined Contributions
NGA	Negotiated Greenhouse Agreements (New Zealand)
NZ	New Zealand
OPC	Ordinary Portland Cement
PAT	Performance, Achieve and Trade Scheme (India)
PMR	Partnership for Market Readiness
PPC	Portland Pozzolana Cement
PSC	Portland Slag Cement
QA	Quality Assurance
RPO	Renewable Purchase Obligations
SA	South Africa
S-CP	Scaled-up Crediting Program
SEC	Specific Energy Consumption
UNFCCC	United Nations Framework Convention on Climate Change

EXECUTIVE SUMMARY

The past year has seen a significant increase in global momentum for climate action. One hundred thirty-seven Parties to the United Nations Framework Convention on Climate Change (UNFCCC) have already submitted their first nationally determined contributions (NDCs) as part of their commitments to the Paris Climate Agreement.[1] With the entry into force of the Agreement on 4th November 2016, and the 22nd session of the Conference of the Parties (COP 22) ending on the high note of further raising ambition, the call to implement these domestically defined commitments has intensified.

Climate policy instruments are increasingly being used or considered by countries to contribute to mitigation commitments. Climate policy instruments, including emission trading schemes (ETS) and carbon taxes (CT) cover about 13 percent of global greenhouse gas (GHG) emissions—a three-fold increase from the past decade.[2]

Benchmarks have been used in climate policy instruments to set targets and thresholds for environmental performance, and to determine the distribution of instrument benefits and obligations. Jurisdictions with mature ETSs, such as the European Union, New Zealand, Tokyo, and California, have been using benchmarks for allocation of emissions allowances in many or all of the sectors that are covered. In recent years, countries developing ETSs have also been exploring the use of benchmarks. For example, South Korea's national ETS uses a benchmarking approach for three sectors. Countries are also showing interest in using benchmarks within carbon tax policies. For example, in South Africa's future CT, sectoral benchmarks will be used to define the level of rebates for covered entities. Furthermore, benchmarks are also being discussed in baseline settings for sectoral crediting programs.

This "Guide to Greenhouse Gas Benchmarking for Climate Policy Instruments" (hereafter, "the guide") is intended to provide policymakers with structured guidance on the development of benchmarks. Practitioners who have already identified the need for benchmarks and are beginning to design them will benefit from the step-by-step approach provided here. The guide draws on over a decade of global experiences in benchmark development, covering practices in 16 jurisdictions that are already using or are in the process of developing a benchmarking approach. While experiences and circumstances of each country are unique, the guide synthesizes these experiences and systematically presents the common practices of countries together with the take-away points of value to a practitioner that is in the process of developing benchmarks. A detailed introduction to the topic, including the basic benchmarking concepts and guiding principles for benchmark development, is also provided.

Key Concepts and Guiding Principles for Benchmarking

This guide introduces the key concepts concerning the use of benchmarks for climate policy instruments. It also explains guiding principles for the development of benchmarking approaches. These aspects are summarized below.

What are benchmarks for climate policy instruments?

A benchmark is a standard of performance that represents the impact associated with each unit of a particular activity. From a climate policy perspective, the impacts could be measured by GHG emissions or energy use, and the activities associated with these can be process outputs (such as products manufactured/services provided or heat produced) or process inputs (such as fuel or electricity consumed). Benchmarks used in climate policies are typically indicators of environmental performance that can be calculated using the following formula:

$$Environmental\ performance = \frac{impact\ (GHG\ or\ CO2\ emissions, energy\ use, etc.)}{activity\ (units\ of\ output\ (product, heat, service)\ produced\ or\ inputs\ (fuel, electricity)\ consumed)}.$$

[1] By mid-April 2017.

[2] Estimated in the 2016 edition of World Bank's "State and Trends of Carbon Markets" report (World Bank, 2016).

Benchmarks can be used when comparing peers against each other or against a certain reference level, such as best available technology (BAT). By setting a common basis for comparison through benchmarks, entities are treated in a similar way under the rules of a policy instrument.

How and why benchmarks are used in climate policy?

Benchmarks can be used in climate policy instruments to set targets or credit thresholds, or as a performance-based approach to distributing instrument benefits or obligations.

In **ETSs**, a benchmark-based approach is one possible method for distributing allowances to entities undertaking a similar activity. All eligible entities would receive allowances corresponding to the amount that would be allocated to a peer performing at the benchmark level. This type of uniform and harmonized allocation treats all entities on the same basis and therefore helps to reduce market distortions that might otherwise arise as a result of the allocation system. Of the jurisdictions surveyed, this approach is used in California, China (Shenzhen ETS), Tokyo, Kazakhstan, New Zealand, and Korea.

In **CTs**, policymakers can provide additional incentives for environmental improvement through tax benefits such as rebates or tax free thresholds, which can be designed in such a way that the overall carbon price signal is preserved. An entity's performance, relative to a benchmark, can be used to determine the level of benefit received. Of the jurisdictions surveyed, this approach is used in South Africa.

In **S-CPs**, benchmarks are used to set baselines or crediting thresholds. These thresholds can be used as a basis to determine and allocate volumes of emission reduction credits generated by participants. Of the jurisdictions surveyed, this approach is used in Columbia and Japan (Joint Crediting Mechanism).

Guiding principles for developing benchmarks

Table 1 outlines the key principles that can be useful when designing a benchmarking approach.

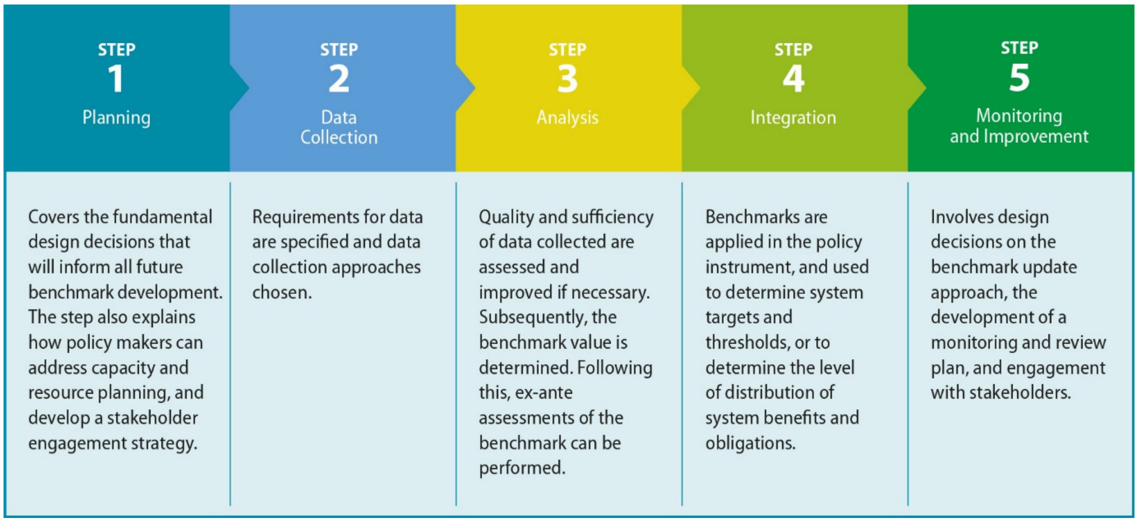
Table 1: Key principles for developing benchmarks

Principle	Description
Alignment with policy objectives	B*enchmark parameters* and *stringency* must be set to align with the policy objectives. The benchmark parameters include the activity covered by the benchmark and the associated impact. The *stringency* is the level at which the benchmark is set relative to the environmental performance of the peers to be covered by that benchmark.
Robustness	Benchmarks must be *accurate, measurable, transparent and relevant* in order to be robust metrics for the performance of entities for which they are applied.
Fairness	Benchmarks must be fair in the sense that they enable a reasonable comparison to be made between an appropriate group of peers and allow for their consistent treatment.
Effectiveness	Where benchmarks seek to incentivize a particular performance, policymakers can look to maximize the effectiveness of this incentive by preferring output to input benchmarks and restricting the differentiation of benchmarks, to the extent possible.
Feasibility	The benchmark approach should aim to achieve robust, fair, and effective benchmarks aligned with policy objectives while taking account of the practical constraints to developing benchmarks.

Steps in benchmark development

The benchmarking process can be condensed into five principal steps, as shown in Figure 1. Each step is presented in turn.

Figure 1: Key steps in the benchmark development



Step One: Planning

The planning step determines all of future benchmark development through fundamental design and implementation decisions.

The first step in benchmark development is planning. This step involves benchmark design decisions, capacity and resource planning, and the development of stakeholder engagement strategy. The key activities in this step are outlined in this section.

Designing the benchmark

Decide which sectors to benchmark. Policymakers should assess the feasibility of developing benchmarks for a particular sector by considering sectoral homogeneity and data availability. Relatively homogenous sectors, such as cement manufacture, can be represented by fewer benchmarks that may be able to be developed more quickly and cheaply. In heterogeneous sectors, such as oil refining and pharmaceuticals, activities performed and outputs are less similar, and more benchmarks may be required. Among the surveyed jurisdictions, oil refining was covered only in California, New Zealand, the European Union, and Kazakhstan. In addition, the availability, quality, and accessibility of data are key factors determining the feasibility of developing a benchmark. The experience of surveyed jurisdictions shows that data issues pose the most pressing challenges at this stage, and the involvement of stakeholders to understand what data is available is essential.

Decide what to benchmark. This is essentially a choice of the impact and activity parameters. The impact parameter is pre-determined by the type of instrument in question, with carbon-based instruments (ETS, CT, S-CP) expressing impact in terms of carbon dioxide equivalence (CO_2e) and energy instruments (energy efficiency trading schemes (EETS) in terms of energy consumed.

Selecting an activity parameter begins by determining which activities are sufficiently similar so that a fair comparison may be made. The tasks involved are outlined below. Since substantial data analysis is required, access to robust and updated data sets on sector activity and environmental impact (verified emissions or energy consumption) are pre-requisites for the analysis. In addition, a deep understanding of products and processes within a sector will be required. Finally, stakeholder engagement is essential to support sectoral understanding and confirm the outcomes of the analysis.

1. **Determine comparable activities for benchmarking.** Economic activities are comparable if they have similar outputs, such as the same product. Therefore, sectoral analysis is necessary to classify products into comparable categories. Such analysis requires an understanding of the activities that lead to the production of the output (defining the "system boundaries" of production) and the environmental impact of those activities. It may be necessary to classify products that are otherwise similar into different categories if differences in inputs and conditions of manufacture that cannot be controlled by the entity then lead to different environmental impacts. However, this should be balanced with practical considerations and the need to preserve the efficiency incentives that a benchmark may seek to provide.[3]
2. **Determine which output benchmarks should be developed.** Having determined the product categories, the next step is to determine whether output benchmarks or alternative approaches should be used. This begins by determining which activities should be covered by the output benchmark. The activities covered should correspond only to activities within the scope of control and responsibility of an entity. In addition, to maximize the emission coverage of an output benchmark, the focus should be on the most common and emission-intensive activities in a sector. Having covered the most emission-intensive activities, the effort associated with covering the remainder may be disproportionate compared to their relative contribution to total sectoral emissions. Alternative approaches, such as fuel, heat, and adapted benchmarks may then be considered. For example, in Kazakhstan output and input based benchmarks were adapted from those used in the European Union.

Choose a methodology to derive the emission intensity of the activities within the benchmark boundary. For instance, in South Africa, the emission intensity of a certain product is defined as the sum of emission intensity of fuel combustion, process emissions, and Scope 2 emissions related to the consumption of electricity. This can be calculated from scratch using data sets or based on pre-existing benchmark values. For example, fuel mix energy intensity benchmarks could already exist. Using existing values can reduce the level of effort required and may be the only option in concentrated economies where the sample size is too small to derive benchmarks that represent best practice. However, existing values should only be used if applicable and they are aligned with the benchmark design choices. International benchmark values often need to be adapted to ensure they reflect local realities. For example, existing benchmarks from the European Union and Australia's Jobs and Competitiveness Program[4] are being used as a reference point for calculating locally appropriate benchmarks in South Africa.

Choose a benchmark stringency level. The choice of stringency level depends on the policy objectives and the intended application of benchmarking within the context of the instrument. It may be decided by taking account of the performance of the peers based on a principle (such as being a certain percentile level derived from peer group performance data), or based on a standard such as the best available level. An average stringency level example is shown in Figure 2, derived by constructing an intensity curve for peer performance data.

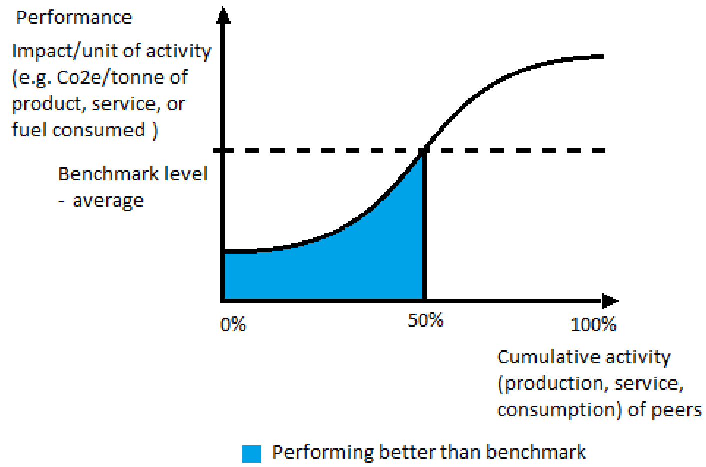
[3] Note that existence and strength of incentives depends on the design of the policy instrument and the application of benchmarks within this context. For example, in S-CP, a clear incentive to perform above the benchmark level is provided if only emission reductions below the threshold are credited, or in CT, if rebates are only given to those performing above the benchmark. In ETS, the use of benchmarks to distribute free allowance may not provide a direct incentive to perform at this level. Instead, an informational signal is provided regarding performance levels in a peer group.

[4] See: http://www.cleanenergyregulator.gov.au/Infohub/Data-and-information/Pages/Jobs-and-competitiveness-program-issued-units.aspx

Figure 2: Average stringency level

Average level: average performance of the selected peers.

Conversely, better-than-average benchmarks would reward only participants performing at or above this level.



Source: Author's illustration

Table 2 shows the stringency of benchmarks used in climate policies in the surveyed jurisdictions. It should be noted that the stringency level relative to the actual performance of peers can only be maintained through regular updates of the data that would reflect their improvements over time.

Table 2: Stringency of benchmarks used in the climate policies in the surveyed jurisdictions

Jurisdiction	Stringency (benchmark level)
Australia (Safeguarding mechanism)	Best practice: weighted average of 10th percentile (proposed)
California (ETS)	90 percent of average or best in class
EU (ETS)	Based on the average of the 10 percent most efficient installations in a sector/subsector in the years 2007–2008
India (EETS)	Best performing plant
Japan (S-CP)	Most efficient under current practices
Kazakhstan (ETS)	Average performance
New Zealand (CT)	10th percentile of international performance
New Zealand (ETS)	Average performance
Tokyo (ETS)	Average performance

Choose a representative historical baseline period. Historical data can be a good predictor of an entity's current and future environmental performance. However, data should be collected from the historical period deemed the most representative of an entity's performance going forward. Collecting data over longer periods (i.e., over three years) maybe costlier, but may be more representative than shorter periods, which may be influenced by unrepresentative shocks. Using data from the years closest to the introduction of the instrument ensures it is most relevant, with an average time span of two to three years typically used (e.g. Australia, California, and India). The choice needs to consider the trade-off between data representativeness (recent years are better) and availability (older data may be more readily available). It is suggested that such decisions be taken in consultation with the benchmarked entities themselves as they are best placed to provide information regarding specific circumstances of each sector.

Creating an enabling environment for benchmark development

Develop a resourcing plan. Experience shows that resource limitations can be a challenge for benchmark development, so careful resource planning is required. Benchmarking is a highly technical exercise, and therefore a dedicated technical team is required for the duration of the exercise.

Policymakers will make key decisions on benchmark design, while a specialist team will need to assess and understand specific technical issues, and manage the extensive stakeholder engagement and data collection exercises. Regarding financial resources, most jurisdictions meet the financial costs of benchmarking through their public budgets. The key cost components are baseline public service expenditure (employees and administrative costs), data collection costs, and the costs of engaging external experts. The timeline for a typical benchmarking exercise for climate policy instruments may vary widely for different jurisdictions and would depend on national circumstances. Experience in jurisdictions (e.g. California and Tokyo, as elaborated in Table 10) shows that the planning and development processes for benchmarking can take three to four years, before benchmarks are available for implementation in an instrument. However, the time and resources needed for different steps would depend on country contexts.

Develop a stakeholder engagement strategy. Early and continuous engagement with stakeholders is not only good practice, but also fundamental for the successful completion of the benchmark development exercise. Experience shows this is also a key challenge. The strategy should outline why, who, and how to engage. Regarding "why," stakeholders will need to be consulted on the design of the benchmark and data availability (covered in Step 1), support implementation (Steps 2–4), and facilitate further improvement to benchmarks (Step 5). As for "who" to engage, this will range from the benchmarked entities themselves (and others affected by the benchmarking) to experts and stakeholders who are considered relevant for implementation and outreach. Numerous approaches exist for how to engage—from targeted engagement approaches (questionnaires were used in India and working groups in South Africa) to public consultations (used in European Union) and online consultations (in Australia). Overall, the critical element for stakeholder engagement is to balance comprehensiveness and transparency with efficiency in decision-making.

Create institutional and legal capacity. Relevant authorities responsible for the design and implementation of the benchmarking exercise must have the institutional capacity to perform this role, meaning the resources and mandate to carry it out. The mandate may be established by memorandums of understanding or contracts between government and entities, or legal provisions may be needed, particularly around data collection, reporting, and monitoring aspects (e.g., European Union and California). While legal provisions can be critical for enabling the exercise they consume a lot of time during the planning stage and may need to be synchronized with the wider legislative planning for the climate policy instrument.

Step Two: Data Collection

Data collection determines the feasibility of benchmark development and is necessary to underpin a robust benchmark.

This step begins by specifying data requirements and, subsequently, choosing and implementing data collection approaches. At each of these stages stakeholder inputs will be important. The key activities in this step are outlined in this section.

Specify data requirements. This involves specifying the data type and format that will be requested from stakeholders to calculate the impact and activity parameters for the chosen historical period. For example, to calculate at the total cement production in India, policymakers specified that data should cover these sub-activities:

- Total cement produced for each grade;
- Total clinker production; and
- Details of additives used.

To arrive at the energy consumption, policymakers requested three-year average data on fuel consumption (by fuel type) and total electricity consumption and source (e.g., grid-purchased or self-generated).

Choose a data collection approach. Three approaches are considered here:

- Collection of pre-existing data (e.g., California);
- Voluntary collection of new data (e.g., Tunisia, EU ETS phase III and Japan); and

- Mandatory collection of new data (e.g., New Zealand, California, India, and the United Kingdom).

The main distinctions concern the type of data collected and the data provision obligation. Data type can be pre-existing data sets or data specifically collected for the purpose of the benchmarking exercise ("new data"). For pre-existing data sets, data providers can be commercial entities, industry associations, or (less often) the benchmarked entities themselves. For new data, the data providers are either the entities themselves, intermediaries who represent these entities (such as industry associations), or contractors/consultant experts working on behalf of the government or the entities.

The data provision obligation pertains to whether data provision is voluntary or mandatory. Data provision is considered mandatory when an enforceable obligation is placed on data providers. Where a mandatory mechanism is used, the engagement with data providers usually falls within wider instrument compliance processes. If it is not mandatory, it may be then based on bilateral engagements or studies commissioned by the relevant authority.

When choosing between the approaches, policymakers should be aware that they have different implications for data relevance and resource requirements. Data relevance is driven by two factors:

- Compared with pre-existing data sets, new data is generally more relevant than pre-existing data, but may require additional resources from the relevant authority and the data providers;
- Whether the data collected is sufficient and representative.

Mandatory approaches increase the chance of obtaining sufficient representative data. Financial, technical, and human resource requirements are driven primarily by the number of engagements with data providers. This will vary with the scale and scope of the data collection exercise in each jurisdiction. Table 3 summarizes these issues.

Table 3: Summary of data relevance and resources required by approach

Approach	Data relevance	Resource requirements
Approach 1: Collection of pre-existing data sets	**Low**—due to the use of pre-existing data and voluntary data provision	**Low**—due to lower number of engagements with data providers
Approach 2: Voluntary collection of new data	**Medium**—while the use of new data increases relevance, voluntary data provision may reduce response rates	**Low/high**—engagement with most benchmarked entities would lead to high costs, but if this can be intermediated, costs can be mitigated
Approach 3: Mandatory collection of new data	**High**—due to the use of new data, and mandatory data provision.	**Medium/high**—engagement with all benchmarked entities; may be mitigated by the integration of relevant costs (including IT) with other instrument systems

Implement selected data collection approaches. This involves preparing and implementing data collection through engagement with data providers. Preparation begins with an assessment of information technology and human resources required by the data collection approach and identification of resource gaps, and subsequently the specification of data collection templates and submission mechanisms. The final stage is that data quality assurance (QA) requirements are specified—for instance, whether the provider must perform QA or ensure third-party verification.

Engage with stakeholders. Engagement with data providers will be necessary to define, request, and support data provision. Support ranges from written guidance to dedicated helpdesks (e.g., New Zealand, and the United Kingdom) and is worth considering as it increases the likelihood of getting timely, sufficient, and relevant data. In addition, experience shows that appropriately addressing stakeholders' concerns over the confidentiality of sensitive commercial data is a key challenge. It is best practice for the relevant authority to agree an approach with the data provider that addresses these concerns adequately, such as allowing data to be anonymized, restricting access to raw data, or only providing data at a certain level of aggregation.

Step Three: Data Analysis

Data analysis determines whether a robust and effective benchmark can be formulated and designed to treat stakeholders fairly.

In this step, the quality and sufficiency of data collected will be assessed and improved if necessary, and the benchmark value determined. Following this, an assessment of the suitability of the benchmark can be performed prior to actual integration into the policy instrument. The key activities in this step are outlined in this section.

Assess and improve data quality. The quality of the data that has been collected to determine the benchmark should be evaluated and additional data gathered if necessary. The quality of the data comprises its accuracy and relevance.

Accuracy checks include:

- Plausibility checks. Comparing existing data with other data sets and relevant sources of information. For instance, in EU ETS Phase III, checks ensured that the most efficient installations were included.
- Consistency checks. Checking consistency in and among data sets, and ensuring reporting uses the correct units of measurement and baseline period.
- Anomalous data checks. This involves checking for outlying values of data (e.g., too high or too low). In California, staff reviewed anomalous data.

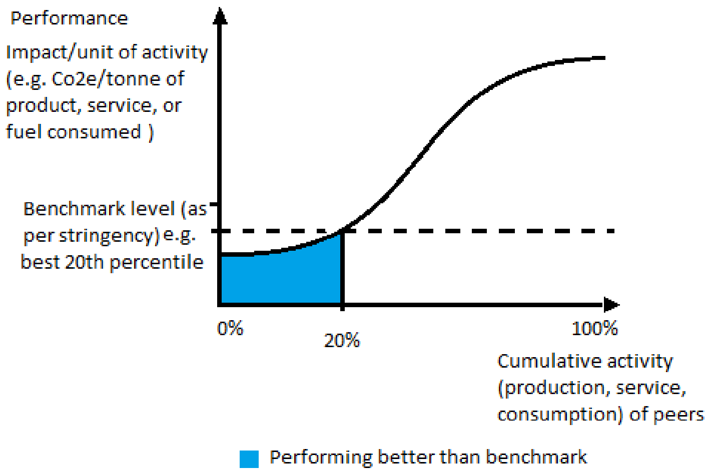
Relevance checks, for consistency with data specifications, regarding:

- Scope (i.e., the sources of emissions covered by the data);
- Historical or baseline time period; and
- Units of measurement—ensure they are known, consistent, and meet the measurability requirement.

Assess and improve data sufficiency. The data collected should be assessed to establish whether there is enough to derive a meaningful benchmark. Data may be insufficient if there are significant gaps. In this case, a possible approach is to estimate/extrapolate based on existing data or focus on data for the facilities whose data is most relevant for the benchmarking exercise (e.g. the top performers, as is proposed in Australia). In this case, one option may be to increase the number of data points using additional years. However, if still insufficient, further data may need to be collected or alternative methodologies used, such as relying on existing international best practice benchmarks.

Determine the benchmark value. The benchmark value is determined by first calculating the emission intensities of benchmarked entities according to the methodology. Using the example of an output-based benchmark, these intensities are then aggregated on an emission intensity curve, and the stringency level (e.g. top 20^{th} percentile) is applied to determine the benchmark value (see Figure 3). At this point the potential benchmarks can be assessed to confirm their suitability. For instance, assessing the benchmark can be either qualitative, by checking whether they are in keeping with the guiding principles, or quantitative, by analyzing the socio-economic impacts of their application.

Figure 3: Example of emissions intensity curve and calculation of benchmark value



Source: Author's illustration

Engage with stakeholders. Stakeholders can be engaged to support the data quality and data sufficiency assessment through a review or audit of the relevant authorities' analyses. In addition, third-party verification may be requested. In Japan's Joint Crediting Mechanism, stakeholders were surveyed for views on methodologies. In addition, stakeholders are naturally interested in the outcome of the data analysis and the final benchmark level. While the final value should take their views into consideration, it should be based on data and consistent treatment in and among peer groups.

Step Four: Integration

Integration refers to the application of the benchmark in the instrument in order to meet policy objectives, and involves ensuring stakeholders' understanding of the benchmark.

In this step, the benchmark values are applied in the context of the policy instrument, determining the level of distribution of system benefits and obligations. Before this can be done, additional activities may be required, as outlined below.

Arrange additional data collection for benchmark application. Relevant authorities may need additional data on the activity parameters to calculate the system benefits or obligations for each entity. For example, when defining first allocation in an ETS, activity parameter data collected in Step 2 (Data Collection) for defining the benchmarks can be used. However, jurisdiction experiences present various situations where fresh activity data collection becomes necessary. These are:

- The facility-level activity data used for calculating distribution levels may be for a longer historical period than used for benchmark determination, to correctly capture fluctuations in production in some industries, for instance economic cycles;
- The data collected during benchmark development either does not cover all installations or was provided in an anonymized manner to the relevant authority;
- The benchmark was based on literature or from benchmark values used in other jurisdictions; or
- The benchmarks are applied to new or modified facilities for which historical baselines do not exist.

For subsequent allocations, new data collection should be undertaken. Policymakers can also include data collection on activity parameters in the monitoring, reporting, and verification (MRV) plan of their instrument.

Consider calibrating distribution levels. Once the final system benefits and obligations are calculated, policymakers might want to adjust these distribution levels further. Jurisdiction practices highlight that adjustments to distribution levels are commonly carried out to implement specific policy goals. For example, in ETSs, policymakers often relax obligations for sectors exposed to carbon leakage[5] by providing a higher share of allowances to them (e.g., in the EU ETS and California ETS). A sector's vulnerability to carbon leakage is defined by its carbon intensity and trade exposure (PMR 2015a). Calibrations may also be carried out to incorporate new entrants, closures, and changes to operations of incumbents. Procedurally, decisions on calibrating the distribution levels are taken at the beginning of the benchmarking exercise or even before—during the design of the climate policy instrument.

Engage with stakeholders. Relevant authorities often provide participants with guidance when implementing benchmarks (such as written guidance and open access tools, and through direct outreach) to acquaint stakeholders with how benchmarks are used in the policy instrument. Experience suggests that transparency in benchmark development, early and adequate engagement with participating entities, and embedding the benchmarking exercise into the instrument's legal framework, can reduce the overall effort required for stakeholder engagement in this step.

Step Five: Monitoring and Improvement

Continuous monitoring and improvement are essential to ensure that benchmarks remain robust, fair, and effective.

This step involves deciding on the benchmark update approach, developing a monitoring and review plan, and in engaging with stakeholders.

Design the benchmark update approach. Updating benchmarks on a regular basis can ensure their continued relevance and stringency. The frequency of updates and the rate of change of the benchmark in each update can be pre-fixed (i.e., **ex-ante approach**). Alternatively, the update can be based on an ex-post review of existing benchmarks with no changes prescribed in advance (e.g., as in California). In the **ex-post approach**, reviews can be pre-planned (e.g., to align with compliance periods). An early decision on benchmark updating, preferably during benchmark planning, is desirable because it sends a clear policy signal to participating sectors.

Decide what circumstances will trigger benchmark updates. Benchmarks reflect the sectoral characteristics of a representative historical baseline period. With the passage of time, the sectoral characteristics change (e.g., efficiency improvements in the sector). Policymakers can define which changes should trigger an update. Revision or change in policy objectives (e.g., increased ambition level) might also require a change to benchmarks.

Develop a monitoring and review plan. After deciding on a benchmark update approach, a plan for monitoring the benchmark's performance and review must be developed. Availability of data is the most critical factor for this step. In some jurisdictions, data reported in the context of other instruments, such as under national reporting requirements (e.g., in Australia), can also prove to be useful. If available, policymakers should check that such data is compatible with their requirements. If required information is unavailable or inaccessible, policymakers need to draft a monitoring and reporting guidance, which outlines clearly the monitoring and verification requirements to the participating entities. It will include information on aspects such as which monitoring variables to report, the acceptable data sources, frequency of reporting, and the verification protocol. Benchmark review can also be included in the overall MRV strategy of the policy instrument.

Engage with stakeholders. Various stakeholders can be actively engaged in the review and update process. This engagement involves communicating the monitoring and review plan to the covered entities, and consulting them on the process. Other relevant stakeholders, such as academic experts, can provide critical insights to the review and update process. Engagements can be structured as formal sectoral working groups and consultations or informal engagement through emails, phone calls, etc. Actively involving participants and sectoral experts in the review and update stage benefits the process by bringing in sectoral expertise and increases stakeholder buy-in for the instrument.

[5] Carbon leakage refers to the risk of entities moving their businesses to jurisdictions with less stringent policies.

In conclusion, this guide draws on the experience gained from jurisdictions worldwide in the development and use of benchmarks for climate policy instruments. The structured guidance on the design, implementation, and improvement of benchmarks provided is summarized in Table 4. Policymakers should note that benchmarking is a resource-intensive and enduring exercise, and should consider whether existing and future resources will be sufficient to undertake the exercise to an adequate standard. For those who do, it is hoped that this guide will be a useful resource.

Table 4: Five steps of benchmark development

	Definition	Key activities
STEP 1 Planning	Covers the fundamental design decisions that will inform all future benchmark development. The step also explains how policy makers can address capacity and resource planning, and develop a stakeholder engagement strategy.	Design the benchmark with stakeholder engagement • Which sectors to benchmark • What to benchmark • How to benchmark Create an enabling environment • Develop a resourcing plan • Stakeholder engagement strategy • Create institutional and legal capacity
STEP 2 Data Collection	Requirements for data are specified and data collection approaches chosen.	• Specify data requirements • Choose a data collection approach • Implement data collection approaches through stakeholder engagement
STEP 3 Analysis	Quality and sufficiency of data collected are assessed and improved if necessary. Subsequently, the benchmark value is determined. Following this, ex-ante assessments of the benchmark can be performed.	• Assess and improve data quality and sufficiency • Determine the benchmark value • Assess the benchmark • Consult stakeholders during assessment
STEP 4 Integration	Benchmarks are applied in the policy instrument, and used to determine system targets and thresholds, or to determine the level of distribution of system benefits and obligations.	• Apply the benchmark in the policy instrument • Communicate with stakeholders and provide guidance on application • Address grievances
STEP 5 Monitoring and Improvement	Involves design decisions on the benchmark update approach, the development of a monitoring and review plan, and engagement with stakeholders.	• Design the benchmark update approach • Develop a monitoring and review plan • Communicate with stakeholders and provide guidance on monitoring and review process

1 Introduction

Policymakers worldwide are increasingly considering using benchmarks to aid in designing climate policy instruments. Benchmarks have been widely utilized when determining allocation of emission allowances in mature emissions trading schemes (ETS), including in the European Union, Quebec, and California. In addition, countries are planning to use benchmarks for their carbon tax policies (e.g., South Africa). Benchmarks are also being discussed in baseline setting for sectoral crediting programs. As countries develop domestic policies to achieve their nationally determined contributions (NDCs) under the Paris Agreement,[6] the uptake of climate policy instruments such as ETS, carbon taxes and scaled-up crediting programs is expected to grow. Benchmarks can play a significant role in ensuring the effective design of these climate policy instruments.

Industry and business have a long history of employing benchmarking techniques to measure and incentivize performance improvement. In a climate policy context, benchmarking can provide policymakers with a granular analysis of the relative environmental performance of covered entities. This information can be used in policy design, especially regarding setting targets or crediting thresholds, and distributing benefits or obligations of different instruments.

This ‘Guide to Benchmarking for Climate Policy Instruments’ (hereafter, the Guide) is focused on the use of benchmarking approaches in the context of specific carbon pricing instruments – namely, **emissions trading schemes** (ETS), **carbon taxes** (CT), and **scaled-up crediting programs** (S-CP). Where relevant, examples from other climate policy instruments, e.g. energy efficiency trading schemes (EETS), with benchmarking applications are included.

The **objectives** of the Guide are the following:

- Assist policymakers in deciding if a benchmarking approach is appropriate for meeting their specific policy objectives;
- Present an overview of benchmark development methodologies and approaches from around the world, including various countries’ experiences with design and implementation; and
- Provide practical guidance to policymakers and practitioners on the main design elements for establishing a benchmark.

To achieve these objectives, the Guide outlines the **guiding principles** and **key approaches** to benchmark development. It then provides **step-by-step guidance** for establishing benchmarks, explaining the key questions and considerations at each stage. The steps focus on the processes for planning, deriving (including data collection), and applying benchmarks, as well as monitoring and evaluation aspects. **Best practices** and **experience gained** from jurisdictions which have implemented climate policy instruments using benchmarking approaches are widely referred throughout the Guide. For each step, empirical evidence is used to determine:

- **Key questions** practitioners should ask when developing the step;
- **Central activities** to undertake, including stakeholder engagement and resource requirements considerations; and
- **Best practices** and **experience gained** from jurisdictions that have undertaken benchmarking activities.

The Guide is aimed primarily at policymakers and practitioners who are developing climate policy instruments, have decided to use benchmarks, and need guidance in benchmark development. Less emphasis is given to the wider upstream policy decisions that might lead to the appropriateness and use of benchmarking. For instance, decisions on the high-level choices between carbon pricing instruments are not covered here. Partnership for Market Readiness (PMR) guidance documents on specific carbon pricing instruments are recommended for this purpose. Issues around carbon leakage and impacts on competitiveness which may lead to cost compensation measures that rely on benchmarks, are discussed only with respect to their role in benchmark development.[7]

The Guide draws on empirical evidence gathered through desk-based research and surveys from 16 jurisdictions. These jurisdictions have been using or plan to use benchmarking, and cover climate policy

[6] The UNFCCC secretariat has established an interim registry for recording the submitted NDCs. Accessible at: http://unfccc.int/focus/ndc_registry/items/9433.php

[7] An elaborate discussion of theory and policy design for carbon leakage can be found in PMR 2015a.

instruments in different sectors. The surveyed jurisdictions, instruments, and sectors covered, are presented in Annex A1. The information presented in this guide was collected through surveys, or through publicly available information, for which links have been provided in Annex A1, complemented by references throughout the Guide.

The Guide is organized around the most important aspects for developing and utilizing benchmarks. *Chapter 0* introduces key concepts and guiding principles for benchmarking. *Chapters 3-7* provide step-by-step guidance in the development of benchmarks for relevant policy applications. Practitioners at any stage of the benchmarking process may use this guidance to support their activities.

The step-by-step guidance is broken down into five main steps and are as follows (Figure 4):

- **Step One: Planning** involves the initial considerations and planning decisions for developing benchmarks. This includes planning the scope of the benchmarking exercise and key parameters which will characterize the benchmark. (*Chapter 3*)
- **Step Two: Data Collection** entails data collection to inform benchmarks. (*Chapter 4*)
- **Step Three: Data Analysis** addresses processes and analytical approaches for using the data to establish values for benchmarks. (*Chapter 5*).
- **Step Four: Integration** involves integration of benchmarks into a policy instrument. (*Chapter 6*)
- **Step Five: Monitoring and Improvement** involves the process of benchmark reviews and updates. These will generally occur after policy instrument implementation, to ensure continued relevance and stringency of benchmarks. (*Chapter 7*)

Figure 4: Overview of benchmark development steps

	Definition	Key activities
STEP 1 Planning	Covers the fundamental design decisions that will inform all future benchmark development. The step also explains how policy makers can address capacity and resource planning, and develop a stakeholder engagement strategy.	Design the benchmark with stakeholder engagement • Which sectors to benchmark • What to benchmark • How to benchmark Create an enabling environment • Develop a resourcing plan • Stakeholder engagement strategy • Create institutional and legal capacity
STEP 2 Data Collection	Requirements for data are specified and data collection approaches chosen.	• Specify data requirements • Choose a data collection approach • Implement data collection approaches through stakeholder engagement
STEP 3 Analysis	Quality and sufficiency of data collected are assessed and improved if necessary. Subsequently, the benchmark value is determined. Following this, ex-ante assessments of the benchmark can be performed.	• Assess and improve data quality and sufficiency • Determine the benchmark value • Assess the benchmark • Consult stakeholders during assessment
STEP 4 Integration	Benchmarks are applied in the policy instrument, and used to determine system targets and thresholds, or to determine the level of distribution of system benefits and obligations.	• Apply the benchmark in the policy instrument • Communicate with stakeholders and provide guidance on application • Address grievances
STEP 5 Monitoring and Improvement	Involves design decisions on the benchmark update approach, the development of a monitoring and review plan, and engagement with stakeholders.	• Design the benchmark update approach • Develop a monitoring and review plan • Communicate with stakeholders and provide guidance on monitoring and review process

2 Key Concepts and Guiding Principles

This section titled *Key Concepts and Guiding Principles* begins by defining benchmarks in the context of climate policy instruments. It then outlines how and why benchmarks are used in the following instruments: emissions trading schemes (ETS), carbon taxes (CT), and scaled-up crediting programs (SC-P), energy efficiency trading schemes (EETS). The section provides an explanation of key concepts such as determining comparable activities and accurately measuring environmental performance in benchmarking. It also presents guiding principles for developing benchmarks for climate policy instruments. *Key Concepts and Guiding Principles* concludes with guidance to aid policymakers in determining if a benchmarking approach is appropriate for their policy instrument.

Key Points

What are benchmarks for climate policy instruments?

- ✓ A benchmark is a standard of performance, representing the impact associated with each unit of a particular activity (for example 780 kg CO2/t clinker).
- ✓ From a climate policy perspective, the impacts could include greenhouse gas emissions or energy use, and the activities can be characterized as process outputs such as products manufactured / services provided, or process inputs such as fuel consumed.

How are benchmarks used in climate policy?

- ✓ Benchmarks can be used within climate policy instruments for setting targets or crediting thresholds, or as a performance-based approach for distributing instrument benefits or obligations.

Key concepts

- ✓ **Determining comparable activities**. Economic activities are comparable if they have the same output, but product quality, processes, inputs, plant age, and location may be valid factors for differentiation.
- ✓ **System boundaries**: Only activities within the scope of control and responsibility of an entity should be within boundaries, and a standardized approach should be applied within a sector. To maximize emission coverage, boundaries should focus on the most common and emission intensive activities in a sector.
- ✓ **Output based benchmarks and alternatives**: An output based benchmark provides a performance standard for the efficiency with which entities convert energy and raw materials into the final output, as measured in terms of the impact parameter. By contrast benchmarks could also represent the impact associated with an intermediary product, such as heat, or an input parameter, such as fuel used.

Guiding principles

- ✓ **Alignment with policy objectives.** There are two important elements of benchmarking that must align with the overall policy objectives: choice of *benchmark parameters* and choice of benchmark *stringency*.
- ✓ **Robustness.** The robustness of the benchmark will depend on *accuracy, measurability, transparency* and *relevance.*
- ✓ **Fairness.** Fairness relates to the concepts of *fair comparison*, and *consistency* of treatment.
- ✓ **Effectiveness.** Given that benchmarks may be used to incentivize entities to perform at the benchmark level, this incentive is best preserved by using output rather than input benchmarks, maximizing emission coverage, and limiting the differentiation of benchmarks.
- ✓ **Feasibility**. Policymakers will need to take a pragmatic approach to balance the need for robust and fair benchmarks aligned with policy objectives and the practical constraints and costs related to defining benchmarks.

Is benchmarking an appropriate choice?

- ✓ Policymakers should be mindful that benchmarking is an enduring process with significant upfront resource requirements, and it may not be feasible in some cases.

2.1 What are benchmarks for climate policy instruments?

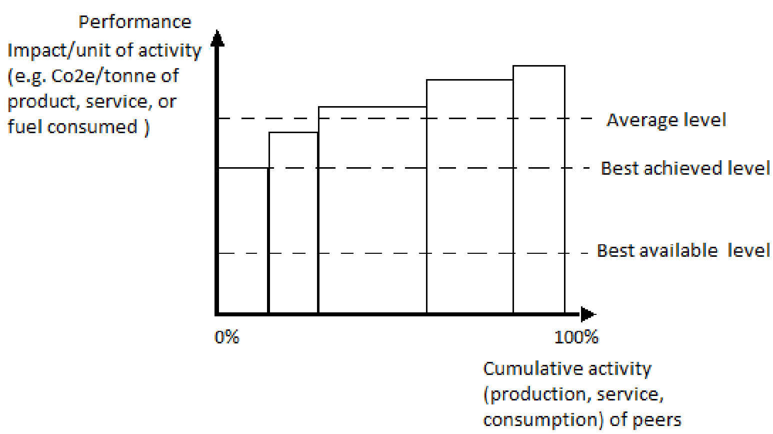
A benchmark is a standard of performance, representing the impact associated with each unit of a particular activity. As a performance indicator, it enables comparative performance analysis. From a climate policy perspective, the impacts could include greenhouse gas emissions or energy use and the activities can be characterized as process outputs (such as products manufactured / heat produced / services provided) or process inputs (such as fuel or electricity consumed). The climate policy benchmark is a measure of environmental performance according to the following formula:

$$Environmental\ performance = \frac{impact\ (GHG\ or\ CO2\ emissions, energy\ use\ etc.)}{activity\ (units\ of\ output\ (product, heat, service)\ produced\ or\ inputs\ (fuel, electricity)\ consumed)}.$$

Benchmarks can be used in performance analysis for climate policy when comparing peers against each other or against a certain reference level such as Best Available Technology (BAT). Setting a common basis for comparison through benchmarks can ensure that entities are treated comparably and fairly under the rules of a policy instrument. In this context, the term peers refer to the entities covered by the policy instrument and which undertake comparable activities.

The relationship between the benchmark and the environmental performance of the peers is shown in Figure 5 below. In this illustration, the peers are ranked from the best performing (i.e., lowest impact per unit of activity) on the left to the worst performing on the right. Possible benchmarks are shown by the horizontal dotted lines.

Figure 5: Illustrative intensity curve used for a benchmarking exercise



Source: Author's illustration

Figure 5 illustrates two important points that are discussed in detail later in this Guide. In particular:

- **Derivation of the benchmark.** The benchmark can be derived directly from the measured environmental performance of the peers, through the construction of an intensity curve. However, other information may be used or taken into account, such as technology standards.
- **Stringency of the benchmark.** The stringency of the benchmark must be defined. This means the level of the benchmark relative to the actual performance of the peers. In Figure 5 three examples are shown corresponding to the peer average level, peer best achieved level, and best available level (i.e., maximum potential given the available technology, even if not applied within the peer group).

2.2 How and why are benchmarks used in climate policy?

Benchmarks can be used within climate policy instruments for setting targets or crediting thresholds, or as a performance-based approach to distributing benefits or obligations. This section explains how and why benchmarks are used in the context of the following climate policy instruments: **emissions trading**

schemes (ETS), **carbon taxes**, **scaled-up crediting programs, and energy efficiency trading schemes (**EETS**).**

Emissions Trading Schemes (ETS). In an ETS, entities are required to acquire and surrender allowances (i.e., the right to emit x tonnes) equal to their verified emissions over each compliance period. The allowances are created by the administering authority and released to the market either for free or at a charge. Allowances may be distributed for free, for instance, to sectors that are at risk of carbon leakage and/or as a transitional approach until all allowances need to be paid for. A method is required to determine the distribution of allowances to entities; one option is to use benchmarks to achieve a harmonized allocation. All eligible entities would therefore receive allowances corresponding to the amount that would be allocated to a peer performing at the benchmark level—importantly, allocations do not take account the individual performance level of each entity.

Carbon Taxes (CT). Carbon taxes apply a carbon price to the emissions of covered entities. In order to incentivize additional environmental improvement, the policy may include some benefit for top environmental performers.[8] This benefit could be designed in many ways, such as a partial or total rebate or the application of a tax-free threshold, and should be mindful of preserving the price signal on emissions. An entity's performance relative to a benchmark can be used to determine the level of benefit received, such that those performing better will receive a greater benefit. In South Africa, the level of tax rebate received depends on an entity's performance relative to the benchmark, and a similar approach was foreseen in the New Zealand Negotiated Greenhouse Agreements (NZ NGAs).

Scaled-up crediting programs (S-CP). Scaled-up crediting programs can be developed either at a sectoral level (sectoral crediting) or at a project/program level (up-scaled project-based crediting). While no implemented examples exist yet, S-CPs are already under discussion in some countries. In an S-CP, the emission reductions achieved against a baseline are issued with credits. Benchmarks can be used to set these baselines and/or to define a performance threshold beyond which credits will be issued. A detailed explanation of the basics of the instrument and use of benchmarking therein is discussed in Box 1.

Although the instruments and applications described above are the main focus of this guide, where applicable other examples of climate policy instruments were considered. This includes the use of benchmarking for target setting in Energy Efficiency Trading Schemes (EETS) such as India's Perform Achieve Trade (PAT) scheme and UK's Climate Change Agreement (CCA) scheme.

Energy efficiency trading systems (EETS). In EETS, an energy savings target is placed on participating entities, who may either comply by undertaking EE measures or by surrendering energy saving certificates, representing verified savings achieved by other participants in the system. Benchmarks can be used to determine an overall peer group performance target. In addition, the overall target can be distributed to each peer based on its performance relative to a peer group benchmark, such that poorer performers contribute to the majority of the target.

Box 1: Benchmarking in scaled-up crediting instruments

Scaled-up crediting is an umbrella concept under which various terminologies and proposals have been developed in the past years. Commonly used terms include sectoral crediting, policy crediting and NAMA crediting. Most of these proposals discuss either up-scaled project-based crediting or sectoral crediting.

Up-scaled project based crediting would focus on types of mitigation activities, for instance renewable energy, fuel, and feedstock switching, within different sectors, such as power generation, industry, and transport. In that respect, it is similar to the Clean Development Mechanism (CDM). It can involve single projects (e.g., energy efficiency improvement in an industrial facility) or a range of similar projects undertaken under an umbrella program (e.g., distribution of energy efficient lighting systems). However, up-scaled crediting would use standardized approaches for setting baseline variables. This standardization could be achieved by developing default factors for emission and fuel characteristics (e.g., grid emission factors) and for performance standards/ benchmarks based on a

[8] In the context of benchmarking for climate policy instruments, environmental performance is measured by a particular environmental impact (Greenhouse gas or CO_2 emissions, energy use etc.) associated with a particular activity (production or outputs or consumption of inputs) performed by an entity in a peer group.

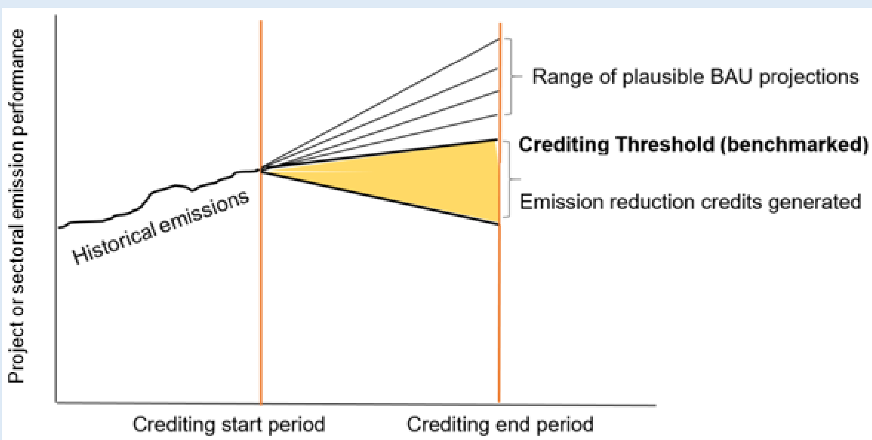
representative peer group on baseline variables. Such standardization approaches are not new and have already been incorporated in many CDM methodologies (Hayashi & Michaelowa 2013).

On the other hand, in **sectoral crediting**, a baseline is set for a broad segment of the economy in the host country (Warnecke et al. 2015). Coverage of a sectoral baseline could be a sector or a sub-sector. Baselines are commonly defined to reflect the business as usual (BAU) situation in a sector. The BAU baseline emissions are assessed against emissions occurring with the mitigation intervention in place to arrive at the emission reductions. Policymakers can issue credits for all the achieved (and verified) emission reductions. Alternatively, ‘crediting thresholds’ can be set to define the level of performance which must be met to issue credits. Figure 6 provides an illustration of crediting thresholds in scaled-up crediting.

Benchmarks can be used to set these crediting thresholds. Benchmarks can be set at average and above average levels of performance for a group of peers in a sector or sub-sector. When set at performance levels that are better than the average sectoral performance, the emission reductions are calculated more conservatively than when using BAU baselines.

Use of benchmarks for setting baselines which apply to the defined peer population can decrease the time and costs borne by participants by avoiding developing baselines on a project-by-project basis, including need for allocating resources towards data collection, monitoring and verification of the baseline. A benchmarked baseline is also generally more conservative as compared to a range of possible BAUs because it is based on the performance of a representative peer group. Hence, benchmarked baselines can reduce the risk of over-crediting or crediting BAU measures. Benchmarked indicators further simplify emission performance estimation in complex systems by aggregating the impact of individual sub-measures, although one must be wary of the risk that higher aggregation increases uncertainty about the achieved reductions. In fact, the extent of aggregation for benchmark setting is a balancing act between uncertainty and simplicity (discussed later in Section 3). In addition, a policymaker must be aware that standardized baseline setting transfers the effort and costs for baseline estimation and update from the participants to the policymakers.

Figure 6: Illustration of crediting thresholds in scaled-up crediting



Source: Author’s modification, based on Prag & Briner 2012, p. 31

Table 5 provides a summary of the application of benchmarking in the policy instruments reviewed. For a summary of how the jurisdictions surveyed in the Guide have applied benchmarking to their climate policy instruments, please refer to Annex A1.

Table 5: Application of benchmarking in reviewed policy instruments

Instrument	Application of benchmarking	Key implications
Emissions Trading Schemes	Method to determine a rules/ approach for the distribution of allowances	Allocations based on peer group standard do not reflect the actual performance of each entity receiving an allocation. Poorer performers will have a greater shortfall compared with the number of allowances they need. The idea behind this is to incentivize improvements in poorer performers and to reward best performers.
Carbon taxes	Method to determine the level of tax benefit (tax rebate or tax free threshold) received.	Entities with better environmental performance receive a greater tax benefit. This reduces the costs faced and provides incentives for environmental improvement, while maintaining a price signal for reductions.
Scaled-up crediting programs	Method to define crediting thresholds	Use of above average stringency generally results in more ambitious thresholds than BAU. Use of thresholds decreases transaction costs for participants, but transfers some of these costs to policymakers.
Energy efficiency trading systems	Method to define an overall target, and/or distribute individual targets relative to the benchmark	Targets are based on granular analysis of actual performance. Poorer performers must contribute most to the target.

2.3 Key concepts in benchmarking for climate policy

A number of key concepts concerning the construction of benchmarks and their update are relevant to the benchmark development and updated process described in this Guide. They are explained below.

Determining comparable activities for benchmarking

One of the greatest challenges in benchmarking is determining which economic activities are sufficiently similar so that they can be considered as comparable and therefore covered by the same benchmark. The optimal use of resources would involve the development of a small number of benchmarks covering a large part of a system's emissions. However, in practice it is desirable that separate benchmarks are used for activities with significant differences, and that benchmarks cover activities over which the emitting entity has control. This is important as the policymaker considers how many benchmarks should apply and what each one will cover.

At a basic level, economic activities are comparable if they have the same objective or output. For this reason, economic activities which have homogenous or interchangeable products or services are usually covered under the same product benchmark. Box 2 explains the meaning of these categories of product type. As a complementary criterion, it may be necessary to determine whether the products being compared are sufficiently similar in terms of their environmental impact for their fair comparison.

Box 2: Type of products

"Towards a more standardized approach to baselines and additionality under the CDM" (Perspectives Climate Change, May 2010) describes these product types:

Homogeneous outputs […] includes products *which are either identical or similar enough that they can be accurately compared without any adjustment. Commodities for example are fully identical products which are solely differentiated by price. This includes […] primary aluminium, drinking water, flat glass, and domestic hot water. Also, most chemical products (e.g., ammonia, methanol, urea, ethylene, hydrogen, oxygen, nitrogen) show either little or no differentiation.*

Interchangeable products*: For many applications, similar products with different properties are found. Although differing properties limit the use of products for certain applications, the room for substitution is extremely large. This possibility for substitution makes the use of a common performance indicator possible and acceptable. This is the case for example with most cement types, which are interchangeable. This might also be the case to some extent for residential units. Also, cooling for residential units with a largely comparable range of cooling temperatures falls into this category.*

However, it is important to carefully consider whether products are genuinely homogenous and interchangeable or whether there are conditions in which one option is preferred. There may be important differences in product quality or properties that lead to materially different environmental impacts of production. For example, while the production of colored or colorless glass in the UK uses the same process, different inputs cause the energy intensity of the production to vary by approximately 90 percent. Cullet (waste glass) can only be used for the production of colored glass and requires less energy to melt and produces less process CO_2. In these cases, separate benchmarks should be considered for products of different quality.

In addition to product quality, there may be other factors for differentiating products, as these factors may not be within the control of the entities covered by the benchmark. In these cases it might not be fair to cover them under the same benchmark. These factors include:

- **Process and inputs**: Electricity production benchmarks could also be differentiated based on inputs (fuel–gas, coal, oil) and processes (technology used, e.g., opened or combined gas turbines). In Europe, there are broadly accepted benchmarks called 'Reference Values' for the efficiency of the electricity generation process for a particular fuel type, as part of the Energy Efficiency Directive.[9] Equally, the environmental impact of the production of crude steel depends significantly on the raw materials that are used. Primary steel making that uses iron ore must be differentiated from secondary steel making that uses scrap metal.
- **Plant age**: as plant efficiency decreases with age, distinguishing between new and existing plants may be an important factor for differentiation. (Perspectives Climate Change, May 2010)
- **Location**: Geographic area may be an important factor for differentiation. For instance, access to fuel types may vary by location and significantly affect the emissions of each entity; for example, emissions for those reliant on fossil fuels may be twice that for those with access to the gas network.

The policymaker must exercise judgment as to whether the factors above are suitable grounds for differentiation. The development of a large number of differentiated benchmarks is resource intensive and it also has important impacts on the incentives that may be created by the climate policy.[10] For example, if a specific benchmark is developed for a small proportion of manufacturers using a less efficient technology, they will have less incentive to invest in more efficient technology. Similarly, certain jurisdictions may choose not to differentiate benchmarks where new projects have wide technology options for investment.

[9] EU Energy Efficiency Directive - https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive

[10] Note that existence and strength of incentives depends on the design of the policy instrument and the application of benchmarks within this context. For example, in S-CP a clear incentive to perform above the benchmark level is provided if only emission reductions below the threshold are credited; or in CT, if rebates are only given to those performing above the benchmark. In ETS, the use of benchmarks to distribute free allowance may not provide a direct incentive to perform at this level; rather, an informational signal is provided regarding performance levels in a peer group.

System boundaries

The system boundaries define the set of activities of which the environmental impact is being considered when developing a benchmark.

As a first principle, only the activities considered within the scope of control and responsibility of an entity should be included within the boundary of their benchmark. For example, electricity generators may not have control over the distribution network, therefore a benchmark based on the emissions intensity of electricity generated at the power station would be a fairer measure of their performance than measuring the emission intensity of electricity supplied to end consumers. Policy objectives also determine the scope of an entity's responsibility by defining the scope of emissions to cover (e.g., direct or indirect emissions), and the treatment of imports and exports to the production process. Importantly, a consistent approach should be applied within and amongst peer groups.

With this in mind, policymakers seek to set system boundaries as wide as possible to maximize emission coverage. An efficient approach to achieving this may be to focus on covering activities common to the majority of entities throughout a sector. For example, a benchmark for the steel sector can be restricted to cover the production of crude steel instead of covering different types of downstream products manufactured on-site from crude steel. Another example is for the glass sector which given in Figure 7. Producing container glass would involve the following steps:

Figure 7: Processes in the production of packed container glass

The emissions from the production of container glass arise in processes 1-4 and 5. The policymaker has a choice whether to measure the activity as units of glass produced at Output 1 stage (container glass leaving the Lehr) or at Output 2 stage (packed glass). A policymaker may decide to restrict the boundary of this benchmark to Output 1, because this covers the most significant portion of emissions, and is common to most glass producers in their jurisdiction. In this case, focusing on a subsection of the production chain (processes common to the majority of entities or the most emission intensive processes), maximizes emission coverage, while using resources and data efficiently. Alternative approaches (discussed below) can be used to cover the emissions arising from Process 5. For example, in the EU ETS the boundary for the container glass benchmark is set at Output 2, whereas due to the practicalities of measuring production the boundary for flat glass (e.g., for windows) is set at Output 1.

As can be seen in the example above, the production of packed container glass can be broken into two parts, and an intermediary product (mass of glass) can be identified. Such disaggregation of production chains, identification of intermediary products, imports and exports, is usually necessary when setting system boundaries.

In addition, in some industrial estates, entities with a centralized boiler may be providing heat and steam for their own use, for the use of other entities, or both. This raises the issue of how to deal with cross-boundary heat which is exported, as arguably, the entities consuming this heat should be made responsible for its environmental impact. This requires a standard methodology to allocate this heat-related impact fairly amongst consuming participants, which involves close examination of cross boundary heat flows.

Output based benchmarks and alternative approaches

An output based benchmark can be understood as a performance standard for the efficiency with which an entity converts energy and raw materials into the final specified output, as measured in terms of the impact of the energy and raw materials. The incentives which may be created for an entity to meet these efficiency standards only relate to the activities within the benchmark system boundary.[11] Therefore, in the example of the producer of packed container glass above, establishing the benchmark

[11] Note that existence and strength of incentives depends on the design of the policy instrument and the application of benchmarks within this context. For example, in S-CP a clear incentive to perform above the benchmark level is provided if only emission reductions below the threshold are credited; or in CT, if rebates are only given to those performing above the benchmark. In ETS, the use of benchmarks to distribute free allowance may not provide a direct incentive to perform at this level; rather, an informational signal is provided regarding performance levels in a peer group.

boundary at Output 1 establishes an incentive for efficiency conversion in Processes 1-4, but not Process 5.

As indicated, alternative approaches exist for covering the emissions related to Process 5, namely fuel benchmarks.

- ✓ *Fuel benchmarks*: Fuels are inputs to the production process expressed as emissions per unit of fuel energy consumed. Fuel benchmarks concern the emissions that arise as a result of the fuel consumed, and can be derived for a single reference fuel or for an assumed fuel mix.

However, the main disadvantage of using a fuel benchmark (e.g., to cover the emissions associated with Process 5 in the example above) is that this only incentivizes a choice of a low emission fuel mix. It misses the opportunity to incentivize an efficient conversion of fuel into the final output.

Nonetheless, as discussed above, it may be desirable to limit the boundary of a benchmark, for instance to increase emission coverage. In addition, it may not be practical to define output benchmarks in complex and heterogeneous sectors with multiple different outputs, particularly if only a limited number of entities are covered by each.

Finally, heat benchmarks are also a commonly used alternative approach:

- ✓ *Heat benchmarks*: Heat is unique because it can be considered an intermediary output of a production process, distinct from the physical product of that process. Heat benchmarks concern the efficiency with which heat is produced and supplied for final consumption, where efficiency can be defined in terms of energy efficiency or emission intensity.

As an intermediary product, the use of heat benchmarks does create an incentive for the efficient conversion of fuel and raw materials into heat, and in this sense may be considered superior to fuel benchmarks. Nonetheless, both these alternative approaches would have to be complemented by an approach to account for process emissions, which are not included.

The majority of the jurisdictions surveyed in the production of this Guide use a combination of output and alternative approaches, as shown in Table 6.

Table 6: Benchmarking approaches in surveyed jurisdictions

Jurisdiction	Output-based benchmarks	Alternative benchmark approaches
Australia (Safeguard Mechanism)	Product-based benchmark or services (transport) (proposed)	Reserve approach (proposed)
California (ETS)	Product-based Benchmark	Historical fuel use-based benchmark
EU (ETS)	Product-based Benchmark	Input-based approaches (as fall back)
India (EETS)	Product-based Benchmark	
Japan Joint Crediting Mechanism (JCM) (S-CP)	Service-based benchmark (street LED lightings in Cambodia)	Process-based benchmark (Centrifugal chillers in Bangladesh)
Kazakhstan (ETS)	Product-based Benchmark	Input-based and adapted approaches (as fall back)
New Zealand (CT)	Product-based benchmarks (oil refinery industry)	
New Zealand (ETS)	Product-based Benchmark incorporating indirect electricity emissions	
South Africa (CT)	Product-based Benchmark (clay, sugar, cement).	Input-based approaches (as fall back) - fuel or electricity/consumption
Tokyo (ETS)	Service based benchmark (Floor area benchmark)	Input based approaches (as fall back) - fuel, heat, and electricity

Updating benchmarks

Regular review and recalculation of benchmark value is necessary to ensure their continued relevance for a sector. This **dynamic benchmarking** approach is work intensive but necessary to ensure that benchmarks remain up-to-date and stringent over time (Warnecke et al. 2015), thus maintaining environmental integrity of the policy instrument. The alternative, which would be not updating the benchmark over time **(fixed benchmarking),** bears the risk that benchmarks become outdated and cease to serve their purpose.

Dynamic benchmarks can be based on **ex-post reviews** of performance of existing benchmarks and sectoral change during scheme implementation. The review informs a policymakers' decision to update the benchmarks, and if updated, what the new values should be (e.g., in UK's CCA scheme).

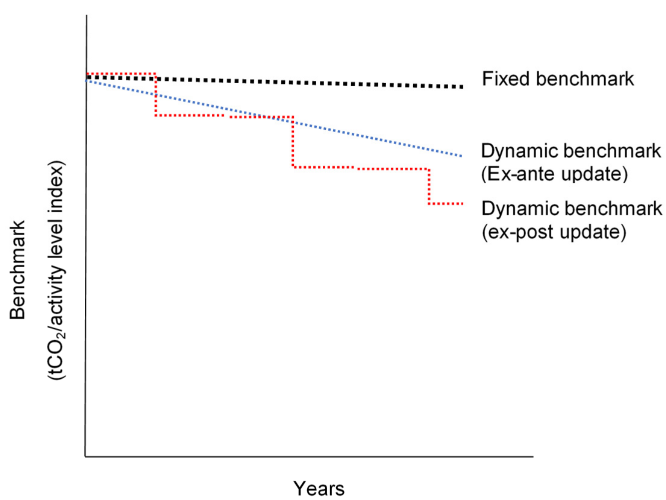
Alternatively, the frequency of update and the rate of change of the benchmark in each update can be tentatively fixed ex-ante (i.e., **ex-ante updates).** Ex-ante approaches are aimed at pushing sectors to improve faster by pre-defining improvement timeframes. Implemented examples are scarce, however, this approach has been proposed by the EU for the fourth phase of EU ETS. The EU Commission's legislative proposal suggests decreasing the benchmark values by the equivalent of a default 1 percent every year, with the updates applying at five-yearly intervals. However, verified annual sectoral improvements will be determined and any sectors improving by more than 1.5 percent will be submitted to a 1.5 percent annual benchmark reduction, and any improving by less than 0.5 percent would lead to a 0.5 percent annual reduction of the benchmark value. The outcome of this proposal is subject to ongoing negotiations.

The difference between the two approaches is only in terms of *when* a policymaker makes the decision to update the benchmarks.

The decision of adopting a dynamic benchmarking approach must be made early in the benchmarking process to avoid policy uncertainty for businesses and signal to the participating sectors a performance improvement roadmap. The procedural details of benchmark update are discussed in detail Section 7 on Monitoring and Improvement of benchmarks.

Figure 8 illustrates the concept of dynamic benchmarking.

Figure 8: Conceptual presentation of fixed and dynamic benchmarks



Source: Author's illustration based on Prag & Briner 2012

2.4 Guiding principles for development of benchmarks

The adoption of a clear set of principles will help the policymaker to develop benchmarks for climate policy instruments. Below are four common principles that should be considered.

Alignment with policy objectives. Benchmarking involves the development of a performance standard that is applied in the policy instrument. There are two important elements of benchmarking that must align with the overall policy objectives: choice of *benchmark parameters* and choice of benchmark *stringency*.

The impact parameter is determined by whether the instrument targets carbon (ETS, CT, S-CP) or energy (EETS) savings, with the metric being emissions or energy respectively. The activity parameter depends on whether output or inputs based benchmarks are desired, so could be denominated as units of production (output based), heat use or fuel use (input based). Table 6 provides examples of the choices made by jurisdictions surveyed, most of which have output and input based benchmarks.

The *stringency* is the level at which the benchmark is set relative to the environmental performance of the peers to be covered by that benchmark. It may also refer to the level of performance expected in the future, when new technologies are adopted, or best available technologies, which may not exist within the peer group.

In ETS, CT, and EETS, where benchmarks are used to set targets or distribute benefits based on performance, the choice of stringency level is a policy decision. This decision is made by assessing the overall impact that stringency level and other instrument design choices (e.g., adjustments to allocation in the case of ETS) will have on entities.

For scaled-up crediting programs, the stringency level is chosen in a way that preserves the environmental integrity of the instrument without making it unattractive to participants (further explained in Section 2.2). Environmental integrity of a crediting instrument is ensured by checking against possible overestimation of credits, usually due to incorrect or less stringent baseline setting. The overestimation risk can be reduced if baselines are set conservatively based on good performers in a peer group.

Robustness. The robustness of the benchmark will depend on *accuracy, measurability, transparency,* and *relevance*.

Accuracy. Benchmarks should accurately represent the total environmental impact of activities within a particular boundary. Each benchmarked activity is comprised of a chain of inputs, processes, and outputs; policymakers must define what activities fall within that boundary.

Measurability. The benchmark parameters must be measurable. Output measures that correspond to the products or services sold will be more routinely measured by peers and therefore more easily covered by a benchmark. Fuel use, which is paid for directly or monitored for environmental purposes will also be measured. Intermediary products, including heat, may not be measured as commonly.

Transparency. As far as possible the data upon which the benchmark is based should be made publically available and subject to scrutiny. Confidentiality concerns may restrict the ability to publish sensitive commercial data; depending on agreement with relevant stakeholders, stakeholder activity data may be published at an aggregated level. The benchmark calculation methods, and benchmark values however, should be transparent and available for scrutiny.

Relevance. Relevance has many different aspects. Regarding data, a benchmark will be relevant if it is based on data from a representative sample of the peer group for the jurisdiction in question. This issue is particularly important if benchmarks are adapted from other jurisdictions or reference values. The relevance of a benchmark may decrease over time, for instance as technology progresses or as policy objectives change. For this reason, it is noteworthy that to remain relevant and timely, benchmarks need to be updated as necessary.

Fairness. Fairness relates to the concepts of *fair comparison*, and *consistency of treatment*.

Fair comparison. As explained in the previous section, one of the greatest challenges in benchmarking is determining what activities are *sufficiently similar* so that a fair comparison may be made. Activities are comparable if they have the same or similar output or have substitutable outputs. However, it may be necessary to differentiate activities according to other factors mentioned above. These include the product qualities, their inputs and processes, and the age and location of the benchmarked plant.

Consistency of treatment. In addition to the principle of fair comparison, a consistent approach is required regarding methodological choices and stakeholder engagement.

- Methodological choices include the definition of system boundaries or the scope of emissions to be covered, and application of a consistent approach within and amongst peer groups. Entities should only be benchmarked on activities considered within the scope of their control. Where there are issues of cross-boundary inputs or outputs, such as heat, a consistent approach must be adopted for sharing the responsibility of these inputs amongst concerned entities.

- Consistent treatment is important in stakeholder engagement throughout the benchmarking process. For instance, all peers should be equally consulted during the design of benchmarks and treated equally during implementation (e.g., data should be gathered using the same approach).

Effectiveness. Benchmarks may be used in policy instruments to incentivize an entity to perform at or above the benchmark level (note however that the strength and existence of such an incentive depends on policy design and the application of benchmarks).[12] Output based benchmarks establish a standard of efficiency for the conversion of inputs into a final output. This means that all decisions and activities within the boundary of an output benchmark are incentivized to reach the level of efficiency established by the benchmark. This includes, for example, the choice of fuel mix, the volume of fuel and electricity consumed, the volume of raw materials used, the efficiency of the technology, and the level of waste in the process.

To maximize this efficiency incentive, these choices are encouraged:

- *Preferring output to input benchmarks*. An input benchmark such as a fuel benchmark only incentivizes a choice of low emission fuel mix, missing the opportunity to incentivize an efficient conversion of inputs into outputs. To the extent feasible, output benchmarks should be preferred.
- *Seeking to maximize emission coverage*. Policymakers should be guided by the objective of covering the most emission intensive activities within a sector. As this is a resource intensive process, a balance must be struck between using output benchmarks for the most emission intensive activities and using alternative approaches for the remainder.
- *Limiting differentiation of benchmarks*. While there may be important reasons to differentiate benchmarks for the sake of fair treatment, differentiation to accommodate specific circumstances may diminish the incentive for emission performance improvement.

Feasibility. The production of benchmarks is a resource intensive process – requiring the ability to plan, collect and verify robust data. Therefore, policymakers will need to take a pragmatic approach to balancing the need for robust, fair and effective benchmarks aligned with policy objectives and the practical constraints to developing benchmarks. This may involve developing a limited number of benchmarks, using fall back approaches or taking a phased approach to introducing benchmarks. Decisions on which benchmarks to develop will be guided by available data, and synergies with other statistical data for climate and energy purposes. Alternatively, benchmarks could be adapted from those applied in other jurisdictions.

Table 7 illustrates the application of the above principles in the case of benchmarking for the South African CT.

Table 7: Application of benchmarking principles in South Africa's carbon tax

Principle	**Recommended approach** (Ecofys and The Green House, October 2014)
Aligned with policy objectives	Stringency: average performance, from the baseline period of 2010-2012 Scope 1 and 2 emissions from combustion fuels as well as process emissions. (Scope 1 emissions are direct GHG emissions from sources that are owned or controlled by the entity (e.g., emissions from fuel combustion and industrial processes). Scope 2 emissions are indirect GHG emissions resulting from the generation of electricity, heating and cooling, or steam generated off site but purchased by the entity.
Robust	All benchmarks to be based on physical indicators (production or consumption of products, raw materials, heat and fuel). Approaches have been adapted from the EU

[12] Note that existence and strength of incentives depends on the design of the policy instrument and the application of benchmarks within this context. For example, in S-CP a clear incentive to perform above the benchmark level is provided if only emission reductions below the threshold are credited; or in CT, if rebates are only given to those performing above the benchmark. In ETS, the use of benchmarks to distribute free allowance may not provide a direct incentive to perform at this level; rather, an informational signal is provided regarding performance levels in a peer group.

Principle	Recommended approach (Ecofys and The Green House, October 2014)
	ETS, but South African products were given special consideration, accounting for the structure and performance of the industry.
Fair	Within each economic sector, differentiated benchmarks were recommended to cover distinct sub-products as necessary using the following factors for differentiation: product quality, inputs, and processes. For instance, in the Iron and Steel sector, distinct benchmarks were recommended to Coke, Sinter, Hot metal, EAF (carbon steel and high alloy steel) and hot metal (COREX/ MIDREX).
Effective	Product benchmarks were prioritized where these could cover at least 80 percent of emissions for most sectors studied (iron and steel, ferroalloys, cement, chemicals, pulp and paper). These do not differentiate by technology, fuel mix, size, age, climate, etc.
Feasible	Alternative approaches were recommended where product benchmarks did not apply, in particular for the petroleum sector, as the number of products was so large as to be prohibitive. Process-based benchmarks were used instead. In addition, where emissions were not covered by the product benchmarks, the recommended alternatives are the use of electricity consumption and fuel benchmark or no benchmark (for a limited number of processes). Although recommended approaches were adapted from the EU ETS and the previous Australian Carbon Pollution Reduction scheme (repealed in 2014), benchmark values are not considered representative of average performance in South Africa, and are instead a starting point for stakeholder consultation.

2.5 Choosing whether to benchmark

As mentioned, benchmarks can be used within climate policy instruments for setting targets or crediting thresholds, or as a performance-based approach to distributing instrument benefits or obligations. In addition, while using them for above-mentioned purposes, the policymaker should ensure that the benchmarking approach is appropriate in the context of the wider instrument adopted and sufficient to reach the end goal. Some of the reasons the surveyed jurisdictions chose a specific benchmarking approach are as follows:

- To provide incentives for environmental improvement, through an implicit performance target (EU), including providing incentives to improve towards an internationally competitive standard (NZ NGAs);
- To reward early action (California, SA) and/or best performers (California, EU), by distributing system benefits in a way that rewards best performers over laggards;
- To conform to international common practice (Kazakhstan, Tokyo ETS); and
- As a proposed methodology for establishing baselines for new entrants (Australia)

Policymakers should bear in mind that although benchmarking is a good option for achieving the objectives outlined above, alternatives may need to be considered[13] for reasons of feasibility and/or alignment with national objectives.

Regarding feasibility and practical considerations, as described later in this Guide, benchmarking is an enduring process with significant upfront resources (financial, technical and human) and data requirements. The cost and feasibility of developing benchmarks will differ considerably between sectors; for example, those industrial sectors with fewer and more standardized production processes would be better candidates than those for which production processes and products vary considerably.

In addition, specifically in the case of S-CP, there is a trade-off between administrator and participant costs. While baseline development costs for participants can be substantially reduced once a

[13] In the context of ETS and S-CP, alternative approaches to distribution of system benefits include grandfathering, whereas project based approaches exist for crediting mechanisms.

benchmarked baseline is established, these costs are then borne up-front by the relevant authority who would be responsible for developing the baseline. This makes the mechanism more attractive to participants (especially compared to alternative crediting approaches not based on benchmarks) but less so for administrators, which policymakers should bear in mind when deciding whether to implement this option.

When faced with feasibility (resource or data) challenges, some jurisdictions have adopted a pragmatic approach of phased benchmark development, piloting it in certain key or less complex sectors. This is the case in South Korea, which for Phase I of its ETS only developed benchmarks for the aviation sector (facility services for domestic private aircraft), grey cement clinker, and oil-refinery sectors. Focusing on a subset of sectors limits the costs and resources required, while still allowing for institutional and private sector learning.

Finally, using benchmarks to incentivize a reduction in emissions intensity for a particular sector, may not always be aligned with national objectives. For example, if a country that is targeting increasing non-conventional renewable energy share and at the same time proposes to develop fossil fuel based generation, may not favor benchmarking its grid emission factor (i.e., a performance benchmark). This reflects particular national priorities for a country at the early stages of climate policy development and with a particular priority focus on promoting certain renewable energy technologies.

3 Step One: Planning

Once policymakers have decided to use benchmarks, the next step is to plan for benchmark development and implementation. This section titled *Step One: Planning* outlines fundamental design decisions that will inform the next steps for developing benchmarks. The section also explains how policymakers can address capacity and resource planning, and develop a stakeholder engagement strategy. The key activities and considerations in *Step One: Planning* are presented below as an overview of the chapter.

Key Activities	Key Considerations
Design the benchmark.	
✓ Decide which sectors to benchmark, i.e., electricity, industry, buildings, waste, agriculture, transport.	Deep understanding of products and processes, data on production and emissions, and close stakeholder engagement are crucial for this step.
✓ Decide what to benchmark, i.e., which activities. This involves analysis of economic sectors and stakeholder engagement to determine comparable activities, based on materiality of environmental impact and other factors for differentiation, and setting system boundaries.	Experience gained from surveyed jurisdictions indicate obtaining data is the most pressing challenge at this stage. This may require the investment of resources in stakeholder engagement for data collection or commissioning third-party studies.
✓ Decide how to benchmark. This includes the choice of the environmental performance methodologies, the use of adapted benchmarks, deciding on stringency levels, and defining the historical baseline period.	These choices must be guided by policy objectives. There is a need to balance cost effectiveness with robustness in choosing methodologies for relevant benchmarks. Above all, these choices must be linked to the end goal of the wider climate policy instrument.
Create an enabling environment for benchmark development ✓ Develop a resourcing plan, including human, technical, and financial resources. ✓ Develop a stakeholder engagement strategy that allows for appropriate engagement throughout benchmark design and implementation. ✓ Create institutional and legal capacity by considering arrangements required to establish effective benchmarks.	Experience gained from surveyed jurisdictions show that institutional and resource limitations, legal framework considerations, and stakeholder engagement may present obstacles. It is important to ensure that there are robust plans in place to address resource requirements and create the necessary capacity.

3.1 Designing the benchmark

In order to design a benchmark, key decisions must be made in terms of which sectors to benchmark, what to benchmark, and how to benchmark.

3.1.1 Which sectors to benchmark?

Economic sectors categorize economic actors into peer groups, and these groups can provide a starting point in deciding which sectors to benchmark. Determining specific sectors and peer groups to be covered involves high-level decision making by policymakers.

In ETS, EETS, and CT instruments, the scope may be limited by policy objectives that define which sectors will be recipients of the instrument's benefits and obligations. For instance, benchmarks to implement allocation in an ETS will naturally only be considered for sectors that are to receive such allocation. On the other hand, benchmarks for SC-P instruments are often developed sector by sector.

Decisions around which sectors to benchmark should also be guided by practical considerations, such as the level of complexity and feasibility for developing a benchmark. Box 3 provides further context in relation to specific carbon policy instruments.

Box 3: Benchmarked sectors in ETS, EETS, and CT instruments

In ETS, EETS, and CT instruments, the choice of which subset of sectors to benchmark is guided by policy objectives. These objectives define which sectors will be recipients of the instrument's benefits or obligations.

In some cases, these benefits and obligations are distributed to all system participants. This is the case in India's PAT scheme, where benchmarking was used to distribute system obligations to each participant, and in South Africa's CT, where it will be used to distribute system benefits (tax rebates).

In other cases, benefits and obligations are distributed only to a subset of system participants. In Phase III of the EU ETS, benchmarking was used to distribute system benefits to the manufacturing and aviation sectors but not to the electricity sector (which had no free allocation). In Australia, proposed benchmarks will set baselines for new entrants only (through an approach similar to S-CP). Therefore, they would be chosen based on where new investments are expected to occur, such as in the mining, oil and gas, and transport sectors.

In the NZ ETS, there are strict criteria for industrial activities to be eligible for free allocation. Only emission intensive and trade exposed sectors are eligible. In some cases, this leads to the selection of a small number of participants per sector, and coverage of some uncommon sectors, such as horticultural activities of tomato and cucumber farming. Note however, that this participant-based eligibility approach was feasible due to a relatively small number of participants in these sectors.

Sectoral suitability for benchmarking

A sector's suitability for benchmarking is largely determined by the homogeneity of activities in the sector and practical considerations regarding data feasibility.

Homogeneity of activities. Benchmark development requires the identification of economic activities similar enough to be compared, and defining a boundary for those comparable activities (known as boundary setting). The entities who perform these activities are categorized as being part of the same peer group.

More homogenous sectors can be represented by fewer benchmarks, which can be developed more quickly and cheaply. In some countries, the cement sector is relatively homogenous in terms of the products produced and its technology, and for this reason is usually one of the first sectors to have benchmarks developed (Kazakhstan, EU, South Africa, India, California, Tunisia, and New Zealand). As a further example, under the CDM, benchmarked/standardized baselines for groups of emitters were encouraged for some straightforward activities such as fuel and feedstock switch (e.g., charcoal production), technology switch and energy efficiency improvements (e.g., in industrial sectors); methane destruction (e.g., landfill gas flaring projects); and methane formation avoidance.

In heterogeneous sectors, activities may not be similar enough to be grouped, and developing output based benchmarks is especially challenging because of the diversity of the products made. Oil refining, food and drink, and pharmaceutical sectors are examples of more heterogeneous sectors. Amongst the surveyed jurisdictions, oil refining was covered only in California, New Zealand, EU and Kazakhstan. While California covered the food and drink and pharmaceutical sectors, alternative approaches (fuel-based benchmarks) had to be used.

Data feasibility. The availability, quality, and accessibility of data are very important in determining the feasibility of developing a benchmark, since it is needed to underpin benchmark design decisions. This should be taken into consideration when determining which sectors or activities within a sector to benchmark. At the preliminary stage, a mapping and assessment of possible data providers at a sectoral level should be performed, and stakeholder involvement is essential so they understand what data is available. The data providers may be the benchmarked entities themselves, research institutions, private sector contractors or sectoral associations. Pre-existing robust data collection mechanisms (GHG or energy MRV or data systems) greatly increase data accessibility for a sector. Consideration can also be given to the capacity of entities to become data providers in the future.

A key experience gained amongst surveyed jurisdictions is that data feasibility issues (limited or lack of availability of data, data sharing concerns such as confidentiality) present the most significant challenges at this stage. Nonetheless, such challenges should be considered in the wider context and their role for reaching the end goal of benchmarking. In California, while verified emissions data was available when benchmarks were being developed (due to the mandatory GHG reporting system) verified product data was not. California arranged surveys to collect the product data, which was a substantial process involving a broad range of stakeholders.

Therefore, in order to decide which sectors to benchmark, policymakers should assess the level of difficulty and feasibility of developing benchmarks for a particular sector by considering sectoral homogeneity and data availability. As mentioned in Section 2, if developing benchmarks for all sectors is not feasible due to practical considerations, policymakers could consider phasing the benchmarking exercise. This involves piloting the exercise in a few key sectors which are more suitable, to enable institutional learning and reduce upfront resource requirements.

Use of benchmarks in unconventional sectors

Benchmarking can find applications in sectors other than energy-intensive industries and energy production (typically covered by ETSs).

The residential buildings sub-sector provides a relevant example. Energy efficiency improvements in buildings are considered to have high mitigation potential but face multiple challenges with respect to accounting for the emission reductions achieved. This sub-sector is also quite disaggregated, with smaller reduction potential per household as compared to commercial or institutional buildings. Nonetheless, the Tokyo Cap and Trade (ETS) system has developed an innovative approach to account for and benchmark emissions in public, commercial, and industrial buildings. This is based on energy consumption (fuel, heat, and electricity) per floor area (m^2) of these buildings in Tokyo, and was used to set carbon intensity targets for facilities.

In the CDM, some methodologies have already explored the use of benchmarks for setting baselines for what is termed as ‘whole-house approaches.’ These approaches package individual mitigation measures in a building such as on lighting, insulation, refrigeration etc. in a single assessment method. For example, methodology AM0091, initially developed for a planned city project in Abu Dhabi—the Masdar city—outlines the following steps to determine a benchmark:

- Baseline emissions are calculated for the covered mitigation measures (electricity consumption, fossil fuel consumption, refrigeration, and water heating/cooling) in a building unit in each building unit category for the baseline year.
- A ratio of baseline emissions and the gross floor area is calculated to arrive at the specific emissions.
- The top 20 percent performing housing units are identified among the baseline houses and a benchmark is defined based on the specific emissions of these top performers in each building unit category per year (t $CO_2e/m^2 \cdot yr$).
- Baseline emissions are calculated using this benchmark.

However, such approaches face several challenges. Very few CDM projects were developed using the benchmarking methods compared to other project types (where baseline setting was easier) because of vast data requirements and aggregation problems in these methods. However, attempts have been made to develop these approaches under different instruments. Mexico, for instance, has tried to adapt the whole house approach in its Nationally Appropriate Mitigation Actions (NAMA) housing database. In the Mexican NAMA, benchmarks were set for the total primary energy demand ($CO_{2}e/m^{2} \cdot yr$) and water consumption of buildings in two categories and four climatic zones. Mexico has an interest in generating credits from this NAMA if a favorable market/financing climate develops in future.

3.1.2 What to benchmark?

Once a high-level decision has been made regarding which sectors to include, the next step is to decide what to benchmark within each sector. The decision of what to benchmark is essentially a choice of the parameters for the benchmarking equation mentioned earlier. These are:

$$Environmental\ performance = \frac{impact\text{: }(GHG\ or\ CO2\ emissions, energy\ use\ etc)}{activity\text{: }(units\ of\ output\ (product, heat, service)\ produced\ or\ inputs\ (fuel, electricity)\ consumed)}.$$

The activity parameters may be expressed in units of output produced or input consumed. Selecting an activity parameter first involves deciding which activities are sufficiently similar so that a fair comparison is possible, as described in Section 2.

The impact parameter is pre-determined by the type of instrument in question. Carbon pricing or GHG based instruments (ETS, CT, S-CP) mainly express the impact of emissions produced in terms of CO_2 equivalent (CO_{2e}) emissions (equivalence facilitates the process of referring to more than one type of GHG). An alternative to expressing in CO_2 equivalence is to simply express the level of a particular gas produced (e.g. tonnes of NO_x). Energy- based instruments, such as EETS, express the impact as energy consumed, for example as MWh of electricity or tonnes of fuel.

Choosing an activity parameter

A policymaker should perform the following four tasks to inform the choice of activity parameters:

1. Determine comparable activities for benchmarking

In practice this means identifying economic activities which have similar outputs, namely homogenous or interchangeable products, with the aim of developing one benchmark per product. Sectors must be disaggregated into discrete products, and may include the identification of intermediary products e.g. in the EU ETS separate benchmarks were developed for coke, sinter and pig iron in the Iron and Steel sector. Once these categories have been identified, complementary analysis to determine whether they are sufficiently similar in terms of environmental impact may be required. Some research proposed that products should be considered sufficiently similar if their emissions different by less than 20 percent (Ecofys et al. 2009).

If the difference is outside this range, this may be a reason for differentiating products and determining separate benchmarks. Other factors to consider for differentiation are product quality, processes, inputs, plant age, and location (as per Section 2.3). In the EU ETS, separate benchmarks were created for Basic Oxygen Furnaces (BOF) steel and Electric Arc Furnace (EAF) steel. The processes used were considered to have a material impact on emissions, so they are not sufficiently comparable. While differentiation may be necessary for fair treatment, restricting the number of differentiated benchmarks is not only resource efficient but it affects the incentives which the benchmark may seek to provide.[14] For instance, in the context of an ETS, if BOF steel receives a much larger free allocation than EAF steel, separate signals are being sent regarding the desired level of efficiency.

This task requires deep understanding of different products and processes within a sector. Determining comparable activities also requires substantial data sets on the activities within a sector (processes and products) and their environmental impact (robust data sets of verified emissions or energy consumption). Certain jurisdictions may choose to commission studies to perform such analyses. Box 4 provides examples of this process in some of the surveyed jurisdictions. Note that having access to

---

[14] Note that existence and strength of incentives depends on the design of the policy instrument and the application of benchmarks within this context. For example, in S-CP a clear incentive to perform above the benchmark level is provided if only emission reductions below the threshold are credited; or in CT, if rebates are only given to those performing above the benchmark. In ETS, the use of benchmarks to distribute free allowance may not provide a direct incentive to perform at this level; rather, an informational signal is provided regarding performance levels in a peer group.

such updated and robust data sets is a pre-requisite for performing the planning exercise, as stated in the "*Data feasibility"* portion of Section 3.1.1.

Once the broad product categories have been identified, it is usually necessary to go to a lower level of granularity to identify the products produced by an entity. This process usually requires significant stakeholder engagement, as described in Section 3.2.2.

Box 4: Sectoral disaggregation: identifying discrete product categories

In order to determine the level where benchmarks should be applied, it is important to understand the activities being undertaken in the participating sectors. In India, the EU and California, regulators emphasized the importance of working with sectors to determine clearly defined products:

- EU ETS: A study (Ecofys et al. 2009) was commissioned to determine the most important and obvious products to include. While PRODCOM[15] data was used as starting point, this had to be complemented with other more detailed classifications. There were also meetings with sector associations, where further requests for benchmarks were discussed.
- California: The NAICS[16] codes were used as a primary classifier. When these were not disaggregated enough, the policymakers defined multiple activities under each code. In addition, conversations with sector stakeholders, site visits, and process flow diagrams helped California staff to understand sector activities.
- India's PAT scheme: The design phase included extensive plant level surveys on energy consumption for 13 sectors, of which eight were finalized to be included in the first cycle of the scheme.

California's experience showed that this can be a time-consuming learning exercise for both policymakers and stakeholders. The development and use of process flow diagrams and site visits were recommended. Process flow diagrams include product lines, where and what type of energy is used and metered, and were prepared and shared by stakeholders with California staff as a guide for identifying appropriate products.

Decisions must be taken at this stage on how to treat entities that produce multiple products, and it may be necessary to benchmark at the sub installation level. In these cases, additional care and time is required to disaggregate the different relevant production activities for any one entity, and in such cases more than one benchmark may be needed.

Box 5 illustrates how this process was undertaken in India and the EU.

Box 5: Determining output benchmarks in the cement sector: PAT and the EU ETS

The high degree of heterogeneity in the energy consumption of units in the sectors covered by the PAT scheme led to their further disaggregation into sub-sectors. All covered entities were mandated to provide detailed data on energy consumption, total production and other key parameters. Using this data, disaggregation was done based on input, process, and output characteristics typical to the sector.

Unlike the EU, the cement sector in India is quite heterogeneous, with varying product types and processes. Specifically, the energy intensity of production decreases with increased blending of additives for different types of cement. For that reason, differentiating product categories based on processes which significantly influence energy consumption was considered appropriate. Sectoral disaggregation was done based on the 'major product' manufactured in a unit. The key products outlined were: Ordinary Portland Cement (OPC), Portland Pozzolana Cement (PPC) and Portland Slag Cement (PSC). Further, plants were also classified into wet plants, white plants, clinkerisation plants, and grinding plants based on processes employed. A conversion factor was employed to convert different types of cement products and exported clinker into the equivalent 'major product'

[15] PRODCOM is a EUROSTAT data resource, which provides statistics on the production of manufactured goods. Source: http://ec.europa.eu/eurostat/web/prodcom

[16] The North American Industry Classification System (NAICS) is the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. Source: http://www.census.gov/eos/www/naics/

produced by that unit (MOP India 2015). The sector-wide target was distributed to these sub-categories on a pro-rata basis and each plant was given a specific energy consumption improvement target based on its performance relative to those in that sub-category.

The approach to defining output benchmarks may differ by jurisdictions, and a contrasting approach was taken in the EU ETS. The EU cement sector is one of the most concentrated in the world (Ecofys et al. 2009), and 92 percent of the cement production in Europe was produced by the same process (dry process kilns). Given this homogeneity, it was evaluated whether clinker (as an intermediary product to the cement production) or cement benchmarks should be developed. A clinker benchmark was chosen, otherwise the allocation to installations producing only the intermediate would become very difficult. A single product benchmark for the production of clinker for the whole EU was developed, since differentiation according to technology, time, inputs, or product quality (aesthetics) would contradict the principle of "one product one benchmark."

2. Determine which output benchmarks should be developed.

Having determined which activities should be grouped under a product benchmark, the next step is to determine whether output benchmarks can be developed, or any alternative approaches should be used. This begins by determining the system boundary of the output benchmark (i.e., those activities that should be covered by an output benchmark).

System boundaries are first defined in reference to the activities an entity has control over, policy decisions determining the scope of the emissions to be included within the boundary, and the treatment of imports and exports (such as cross-boundary heat flows). Within these constraints, policymakers should aim to maximize the coverage of emissions by focusing on the most emission intensive activities, and those common to the largest number of entities within a sector. A standardized approach common to all peers within a group is required, and close engagement with benchmarked entities can be useful to understand their scope of control and responsibility in the production chain. Links to further sector specific guidance for establishing organizational boundaries are provided in the footnote.[17]

Having covered the most emission intensive activities, the effort associated with covering the remainder may be disproportionately large given that the additional activities' contribution to total sectoral emissions are usually small or negligible. A pragmatic approach is to aim to cover the majority of a sector's emissions with output benchmarks. For the remaining emissions, and in situations of insufficient data, alternative approaches can be considered, such as fuel or heat benchmarks.

3.1.3 How to benchmark?

This stage defines how to calculate the benchmark value/s. This involves both the choice of methodologies, and four key decisions that will inform the approach used to derive the benchmark values:

- Choice of the methodology for deriving the environmental performance of the activities benchmarked. Policymakers must choose between deriving these from scratch or using reference values;
- Whether to use adapted product benchmarks from other jurisdictions;
- Choice of the benchmark stringency level; and
- Choice of the historical baseline period for the data.

The four key decisions are explored in depth below. The results of these decisions will affect the design of the entire benchmarking process going forward.

Deriving emissions intensities of benchmarked activities

The emission intensity of production for obligated entities is equivalent to the sum of the emission intensity of activities within the benchmark boundary. Considering a simple product benchmark, in a

[17] European Commission Guidance Document No. 9 "Sector specific guidance (GD9) (https://ec.europa.eu/clima/policies/ets/allowances_en#tab-0-1); Aluminium sector by WRI/WBCSD (http://www.world-aluminium.org/media/filer_public/2013/01/15/fl0000127.pdf), and CSI for the cement sector (http://www.wbcsdcement.org/pdf/tf1_co2%20protocol%20v3.pdf)

jurisdiction where electricity emissions are within the boundary, the emission intensity of production can be calculated as demonstrated in Figure 9 below.

Figure 9: Calculating the emission intensity of production

$$\text{Emission intensity of production } (tCO2/t\,product) = \underbrace{\frac{(Volume\ of\ fuel * emission\ factor)}{tonnes\ of\ product}}_{\text{Emission intensity of fuel combustion}} + \underbrace{\frac{(Volume\ of\ process\ emissions * emission\ factor)}{tonnes\ of\ product}}_{\text{Emission intensity of process emissions}} + \underbrace{\frac{(Volume\ of\ electricity\ consumed * emission\ factor)}{tonnes\ of\ product}}_{\text{Emission intensity of electricity emissions}}$$

Source: Author's illustration

To enable accurate calculation of emission intensity, emission and production data must be collected for each peer. Such verified emissions data sets may already exist in some jurisdictions (e.g. collected for the purposes of emissions monitoring), otherwise engagement with entities to collect the data may be necessary. Although not the focus of this Guide, readers should note that substantial guidance exists for quantifying GHG emissions at an entity or facility level, and further links are provided in the footnote.[18]

Production data will also need to be collected, such as product type, and volumes of production by entities in given years. Once all necessary data has been collected, it is then aggregated to form an emission intensity curve. This is a visual representation of the data in which the data points are plotted in increasing intensity. In order to derive the benchmark value, the stringency level must be selected, and the benchmark will be calculated accordingly. For instance, if an average stringency level is chosen, the benchmark value is the average emission intensity of the peer group. Box 6 provides an example of how the emission intensities are calculated in South Africa. This generic formula is used for all benchmarked products and sectors.

Box 6: Calculation of emission intensity in South Africa

The product emission intensities are calculated via the following generic equation:

Ypi= ((FCxi*Xfxi) + (ECxi*Xexi) + PExi) / Pxi

Ypi – GHG emissions intensity (Scope 1 and 2) of the product i covered by a product benchmark in tCO_2e /t product

FCxi – Fuel consumption for the production of product i in the baseline period x in GJ

Xfxi – Measured and verified actual emission intensity of direct fuel use for the production product i in the baseline period in tCO_2e/GJ

ECxi – Electricity consumption for the production of product i in the baseline period x in MWh

Xexi – Measured and verified actual emission intensity of electricity consumption for the production of product i in the baseline period x in tCO_2e/MWh

PExi- Process emissions from the production of product i in the baseline period x in tCO_2e

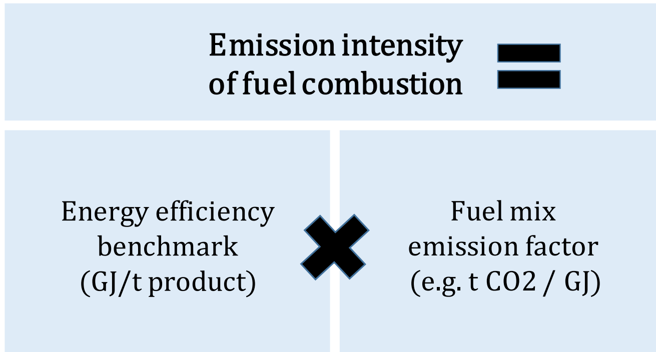
Pxi – Production of product i covered by product benchmark in the baseline period x

[18] The Greenhouse Gas Protocol contains standards and guidance for preparing GHG inventories at corporate and city level, and quantifying the GHG benefits from projects. Sec specific emission calculation tools sets have also been developed. These can be complemented by sector specific guidance, for instance that elaborated for the Aluminium sector by WRI/WBCSD (international Aluminium Institute 2006) and by WBSCD for the cement sector (WBSCD 2011). Full reference details included in the "References" section of this Guide.

STEP 1 Planning
STEP 2 Data Collection
STEP 3 Analysis
STEP 4 Integration
STEP 5 Monitoring and Improvement

In some cases, if jurisdictions have access to existing benchmark values, it may be desirable or necessary to use them. The emission intensity of fuel combustion may be derived from pre-existing benchmarks, such as energy efficiency benchmarks and fuel mix emission factors, as is illustrated in Figure 10.

Figure 10: Emission intensity of fuel combustion



Source: Author’s illustration

Policymakers typically may have access to existing studies defining fuel mix and energy efficiency. For example, existing emission factors for a specific type of fuel mix, such as an average fuel mix for the sector or an environmentally efficient fuel mix, can be used to define benchmarks. In the EU pre-existing benchmark values on best available technology from Best Available Techniques Reference (BREF) documents under the Industrial Emissions Directive were used to define some energy efficiency benchmarks.

However, it is important to consider carefully whether such existing benchmark values are applicable for the policy instrument in question and still reflect the national circumstances. This will be determined by how closely the peer group on which the reference values are based match the group targeted under the policy instrument in question. For instance, reference values developed for the EU may not be appropriate for other jurisdictions. It will also depend on whether the values match the design choices such as the stringency level of the desired benchmark. For instance, the use of “best available” international reference values may not be appropriate if a jurisdiction has chosen an average stringency level.

Australia is proposing to use a reserve approach where there is insufficient data to determine a benchmark value from scratch (Australian Department of Environment 2016). The proposed approach gives an example of one possible way of estimating a benchmark value based on a ‘theoretical’ leading class facility—a hypothetical facility which is assumed to have access to the lowest emissions intensive technologies and practices currently deployed within Australia. The emissions intensity of this hypothetical facility is estimated and used to establish a benchmark value.

Adapting existing product benchmarks from other jurisdictions

An alternative to developing benchmarks from scratch would be to adapt benchmarks from other jurisdictions to the national context. The advantage of adapting existing benchmarks is that it can significantly reduce resources required to benchmark.

However, existing product benchmarks may only be adapted if the products (processes and inputs) are considered sufficiently similar with regards to the environmental impact as that of the jurisdiction in question. In the case of South Africa and Kazakhstan, while production processes and technologies were considered similar to those in the EU, benchmarks had to be adapted to allow for difference in inputs. In Kazakhstan, EU ETS benchmarks were adjusted to account for the level of economic burden and industrial development in Kazakhstan compared to the EU. In South Africa, benchmark values from the EU and Australia’s Jobs and Competitiveness Program are being used as a starting point for stakeholder consultation, and this will act as a reference to take into consideration when calculating local benchmarks.

Small or concentrated economies may be more reliant on adapted benchmarks than others. For instance, the low number of entities (some as low as a single entity) in some sectors in NZ did not allow for the development of meaningful best practice benchmarks from local data, as population size is too small. In the oil refinery sector, the Solomon's Energy Intensity Index[19] for refineries was translated into an annual emissions pathway for NZ refineries. Since the NZ benchmark is judged against *international* best practice, this locally specific benchmark had to be supported by another study to establish the distribution of refinery performance worldwide, and estimate how this benchmark was projected to move over time.

Another approach, particularly useful for S-CP, is to use adapted benchmarks for key parameters required for emission intensity calculation. Many JCM methodologies use default factors for key parameters in the baseline emission calculation. In cases where getting country-level data of best performers is difficult, or the best performance is also not up to the mark compared to international common practice, defaults from Japan or other mechanisms such as CDM are used. For instance, JCM methodology KH_AM001 for 'Installation of LED street lighting system with wireless network control' uses a default value for luminaire efficiency of reference street lighting system taking into account Japan's highway lighting standards for major arterial roads. The baseline emission performance is then calculated on the basis of the rated power consumption of project street lighting systems, ratio of luminaire efficiency of project/baseline lighting, operating hours of baseline lighting systems and CO_2 emission factor of the grid. In other methodologies, default factors designed under CDM have also been used. In a sectoral crediting mechanism, such default parameter values can be used in calculation of the default sectoral performance. However, adoption of these measures could be difficult in practice.

Choosing benchmark stringency levels

Once the emission intensities of a group of peers under the same product benchmark have been estimated, they can be aggregated into a single emission intensity curve. This curve will provide information regarding the relative environmental performance of each peer being benchmarked, and the range of environmental performance within the peer group.

In order to define a benchmark or standard for environmental performance, a stringency level must be applied. This may be determined by taking into account the performance of peers on the intensity curve, (such as being a certain percentile within the peer population) or a standard such as the Best Available Technology (BAT). The choice of stringency level will vary in accordance with the policy objectives and the application of benchmarking within the context of the instrument.

For example, in ETS, CT and EETS, where benchmarks are used to set targets or distribute benefits based on performance, the choice of stringency level is a policy decision which takes into consideration the impact of stringency levels on affected entities. In ETS where free allocation is usually used for sectors at risk of carbon leakage, the chosen level of stringency is part of a wider policy decision that determines level of free allocation these sectors should receive. In some cases, average stringency levels are chosen (Kazakhstan, New Zealand, Tokyo), while others chose better-than-average (EU, California). Similarly, for CT the choice of stringency is a policy decision relating to the level of thresholds and rebates which should apply to particular levels of performance within a sector. These decisions are made by assessing the overall impact which stringency level and other instrument design choices (e.g., adjustments to free allocation in the case of ETS) will have on entities.

In S-CP a key policy objective is to ensure the mechanism delivers actual emission reductions (i.e., has high environmental integrity). In this context, it is important to choose a benchmark level that balances the possibility of delivering excess or fewer credits. An emission intensity threshold set at a lower carbon intensity (i.e., a less stringent benchmark) may be easily met by many entities, hence generating emission reduction credits for activities which could have taken place in a business as usual situation. Alternatively, a benchmark set at a high carbon intensity (i.e., a more stringent benchmark) may not be met by many entities in spite of genuine mitigation efforts. This can deter participation in the mechanism. Defining an appropriate level of stringency on pure technical standpoint has been a difficult and debatable issue in CDM. Due to this reason, most CDM methodologies that use performance benchmarks resort to using the *average of top 20 percent performers* benchmark level, politically agreed upon in paragraph 48 (c) under the Marrakech Accords (Hayashi & Michaelowa 2013). Departures from this have been few in CDM, such as use of the *average of the top five performers* for methodologies

[19] Solomon's Energy Intensity Index is an energy efficiency benchmark value calculated based on world class facilities.

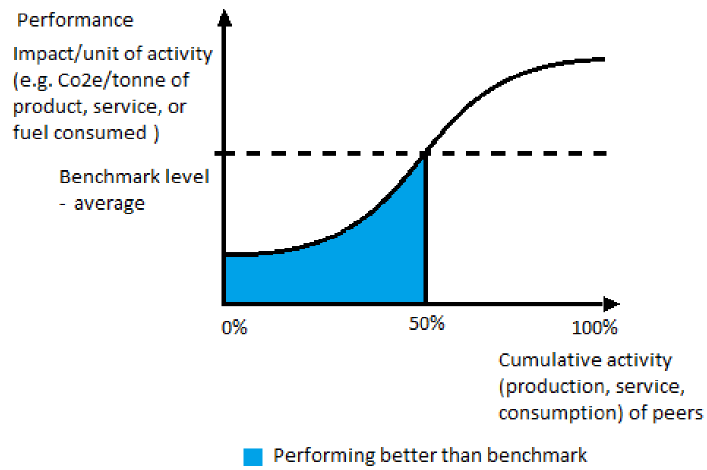
ACM0005 and ACM0015 in the cement sector, and the *average of the top 15 percent performers* in methodology ACM0013 for clean fossil fuel power plants.

The most commonly chosen approaches for deriving benchmark stringency levels are presented in Figure 11.

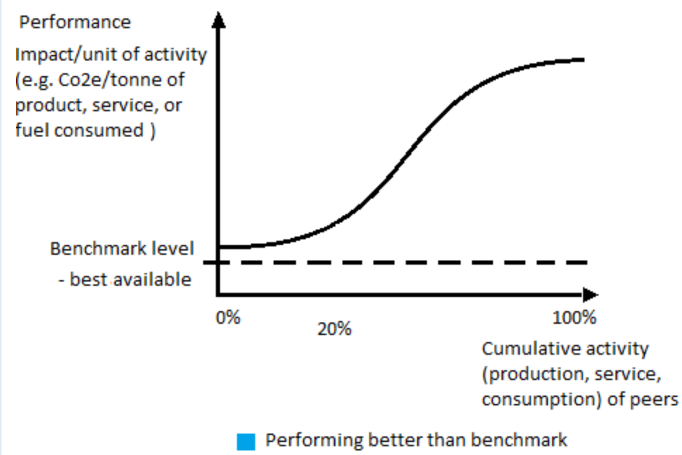
Figure 11: Approaches for deriving benchmark stringency levels

Average level: average performance of the selected peers, a simple exercise.

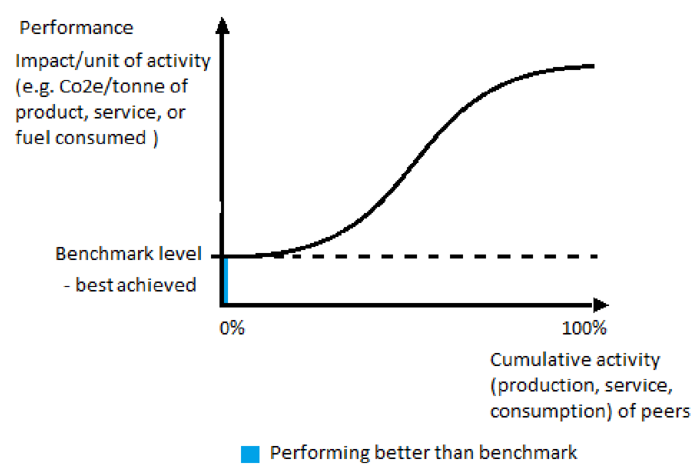
Better-than-average benchmarks, on the other hand, would reward only participants performing at or above this level.



Best available level: the performance of a hypothetical best performer is estimated, assuming the best technology and practices available. This is a complex exercise, reliant on numerous assumptions.

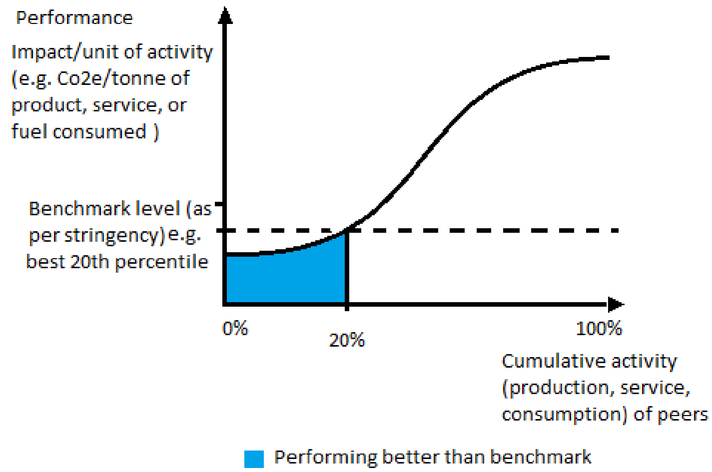


Best achieved level: based on actual performance of the best amongst peers, therefore measurable, also known as the 100 percent percentile.

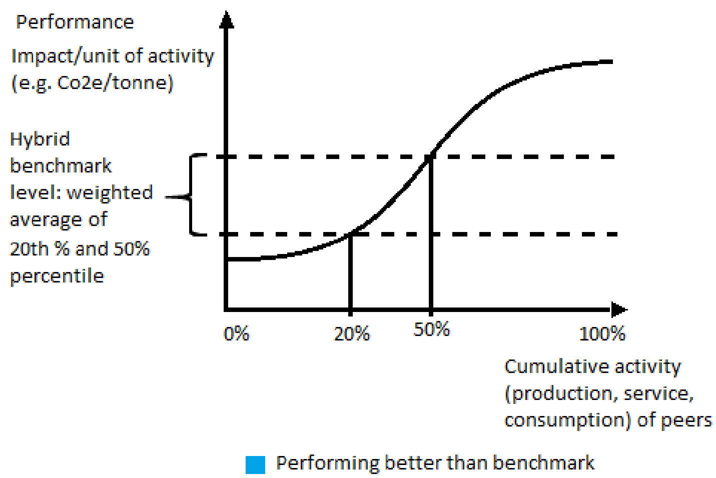


Top percentile levels: based on the cumulative production of a group of peers, targets a particular top percentile of performance – e.g. 20th percentile. This is less stringent than the *best achieved level*, and is considered more balanced.

Under a stringent benchmark relatively few of the most efficient installations would be allocated close to or above their requirements. In particular, in S-CP, participants could be deterred from participating in the mechanism if the benchmark is too strict.



Hybrid models: Whereas the options above only consider a certain level or interval of performance, a hybrid model can bring together more than one interval or level. For instance, this can be achieved by combining a top and bottom percentile, in a weighted average. This would take into account achievable levels, and also the distribution for least efficient peers, but is less stringent overall.



Source: Author's illustrations for all

Table 8 shows the stringency of benchmarks in the surveyed jurisdictions. It should be noted that due to technological progress and competitive pressures, sectors tend to improve their environmental performance as time passes irrespective of such incentives. In other words, an increasing number of participants will be able to reach the absolute benchmark level, whereas in relative terms, the percentile level has decreased. The relative level can only be maintained through regular updates of the data and the benchmark level to reflect technological improvement over time.

Table 8: Stringency levels in surveyed jurisdictions

Jurisdiction	Stringency
Australia (Safeguarding mechanism)	Best practice: weighted average of 10th percentile (proposed)
California (ETS)	90 percent of average or best-in-class
EU (ETS)	Based on the average of the 10 percent most efficient installations in a sector/subsector in the years 2007 - 2008
India (EETS)	Best performing plant
Japan (S-CP)	Most efficient under current practices
Kazakhstan (ETS)	Average performance
New Zealand (CT)	10th percentile of international performance
New Zealand (ETS)	Average performance
Tokyo (ETS)	Average performance of facilities covered in the previous program

Choosing a representative historical baseline period

A benchmark in the context of this Guide is an environmental performance standard for a group of peers at a particular point in time. Historical performance data is gathered because it offers more robust data that is better for comparisons than use of forecasted performance, for example. When choosing the historical period from which to collect data, the objective is to choose a range of historical years that will be representative of an entity's activities going forward. It is suggested that such decisions be made following engagement with the benchmarked entities, to form a view regarding availability of this data and its representativeness over the future years.

It can cost more to use longer (e.g., greater than three years) historical periods to gather data but the data that is gathered may be more representative of average activity going forward. In addition, the use of data from historical years can mitigate the risk that entities manipulate activity during specific years to influence the benchmark level. Conversely, shorter baselines may be influenced by short term or unrepresentative shocks, such as production level variations due to economic downturns. More recent baselines will take better account of progressive efficiency improvements from BAU technological development but may prove more difficult to collect (complete, verified data could become available with a one- or even two-year delay).

As a general rule, years closest to the introduction of the instrument are recommended, using an average time span of two to three years, to avoid distortion by unrepresentative years. Of the 13 surveyed jurisdictions, 10 chose baselines within this timeframe. Table 9 presents the selections of some of the surveyed jurisdictions.

Table 9: Historical baselines periods chosen in selected surveyed jurisdictions

Jurisdiction	Historical baseline period	Date introduced
California (ETS)	Typically, 2008–2010 If those data years were not representative of normal operation years, most representative years were selected.	2011
EU (ETS)	2007 and 2008	2013
India (EETS)	PAT I: three-year average (2007/08 to 2009/10), next cycle onwards baselining will be on a rolling basis (1 year)	2012
Japan (S-CP)	Determined by the Joint Committee between Japan and host countries	n/a
Kazakhstan (ETS)	2010 - 2015	2013
New Zealand (ETS)	Default years - financial years 2006/07, 2007/08 and 2008/09	2010
Tokyo (ETS)	Covered facility data (FY2005-FY2007) in the previous Carbon Reduction Reporting Program, three consecutive years between FY2002 and FY2007 (Existing facilities)	2010

3.2 Creating an enabling environment for benchmark development

In order to successfully design and implement benchmarks, policymakers must create an enabling environment. This includes having access to the right resources (human, technical, and financial). It also involves ensuring there is the institutional and legal capacity for policymakers to perform their role, and develop a stakeholder engagement strategy.

3.2.1 Develop a resourcing plan

The benchmarking exercise requires technical, data, administrative, and legal human resources, as well as financial resources. The experience of surveyed jurisdictions shows that resource limitations (manpower, technical skills, organizational setup, other resource shortages/delays) were a pressing challenge. Therefore, it is important to plan for meeting such requirements in advance.

In relation to human resource requirements, teams with different skill sets—policy making authorities, technical teams, and relevant administrative authorities—are required for different tasks during the process.

Benchmark development is a highly technical exercise, and technical expertise is required in particular during the design and analysis stages. Policymakers can decide whether they undertake the entire benchmarking exercise themselves, or whether they will simply provide a framework in which stakeholders must develop benchmarks, as is the case in South Africa. South Africa has provided a guiding framework which outlines generic principles and recommendations of suitable benchmarking approaches for the regulated sectors. The industry associations are then expected to choose their own consultants to establish the benchmarks. On the other hand, many jurisdictions closely engage in the benchmark development process and use in-house technical experts in addition to receiving inputs from external agencies.

If policymakers are undertaking the exercise themselves, a dedicated *technical team* must be created, consisting of economists, engineers and scientists in different capacities, with good sectoral understanding. These experts can be sourced from within the government if exists (as in the case of India's PAT scheme) or recruited/hired externally. External expertise is also usually hired for preliminary scoping studies or capacity building for the technical team. It is also common practice to bring on board third-party verifiers for auditing and verification of firm level data during data analysis. The same technical team can continue to support and review the scheme implementation (e.g., in Tokyo ETS).

Less technical expertise and more administrative responsibilities are required during the implementation stages. In this Guide the term *relevant authority* indicates the public-sector entity responsible for implementing the decisions of policymakers. Relevant authorities undertake administrative tasks of stakeholder engagement and communication, ranging from one-on-one meetings, workshops, consultations, online inputs, feedback to received inputs, etc. Further resources must be dedicated towards the data collection, reporting, and monitoring. These typically include data management systems such as electronic or manual collection tools and portals (Excel templates, online forms), processing software (database software or more advanced solutions if needed), secure storage and other IT resources (professionals, equipment, etc.). Authorities may already have such resources in place. In all cases, the authority has the option of outsourcing these tasks to third parties.

In relation to the cost implications of benchmarks on final allocations/rebates/reduction targets for the industry, relevant authorities often face extensive lobbying and negotiations with industries trying to influence the process in favor of their sector. Therefore, it becomes important to bring on legal experts as well. Common legal matters include supporting industry negotiations from a legal standpoint, representing the administrator in case of potential legal issues, and advising how to integrate benchmarking into the policy and legal frameworks.

In terms of financial resources, most jurisdictions meet the financial costs of benchmarking through their public budgets. International support such as that provided by the PMR could also be explored to meet part of the costs, especially for external experts supporting benchmark development.

Based on information provided by the surveyed jurisdictions, financial resources required for benchmarking can be grouped into following three categories:

1. **Baseline public service expenditure**. This includes costs related to the personnel employed with the administrator, costs for outreach, etc.
2. **Data costs.** These costs would differ substantially based on the data collection approaches chosen. In cases where no prior data exists, resources might be required to gather data, for example. through voluntary surveys, for making informed decisions regarding the policy instrument. The costs of setting up and maintaining the reporting system under the instrument are also associated with data expenses. This might be particularly intensive and resource heavy for developing countries where national reporting frameworks are not well developed or are non-existent.
3. **Expenditure towards external experts**. These expenditures include consultancy fees for experts engaged in the benchmarking process.

The timeline for a typical benchmarking exercise for climate policy instruments may vary widely for different jurisdictions and would depend on national circumstances (e.g., prior work and data availability on the selected sectors, availability of sectoral expertise, and governance bodies). However, jurisdiction experiences highlight that **the benchmarking exercise can take three to four years of planning and development before the benchmarks are ready for use in an instrument.** Different steps may be more or less time-intensive for the relevant authorities. In general, the *planning* step is most time-intensive while *data analysis* is least time-intensive. *Data collection*, *integration,* and *monitoring and improvement* phases require intermediate time-effort. Similarly, the resources and costs of different steps would differ.

Table 10 provides example of resource requirements and timeframes needed for the Californian ETS and Tokyo ETS.

Table 10: Example of resources required for benchmarking: California ETS and Tokyo ETS

Resource requirements	California ETS	Tokyo ETS
Step One: Planning		
Time (months)	12	36 (also includes the following 3 steps)
Human and technical resources	5 staff (scientists/engineers)	Many Tokyo Metropolitan Government (TMG) staff
Costs	Consultancy fees, personnel	Personnel expenses for TMG staff
Step Two: Data Collection		
Time (months)	6 (including step 3)	As above
Human and technical resources	5 staff (scientists/engineers)	Many TMG staff
Costs	Personnel expenses	Personnel expenses for TMG staff
Step Three: Data Analysis		
Time (months)	6 (including step 2)	As above
Human resources	5 staff (scientists/engineers)	Many TMG staffs and some consultants
Costs	Personnel expenses	Personnel expenses for TMG staff
Step Four: Integration		
Time (months)	24	As above
Human resources	5 staff (scientists/engineers)	Many TMG staff
Costs	Personnel expenses	Personnel expenses for TMG staff
Step Five: Monitoring and Improvement		
Time (months)	ongoing	12
Human resources	5 staff (scientists/engineers)	5 TMG staffs on average
Costs	Personnel	Personnel expenses for TMG staff

Source: Author's own data collection

3.2.2 Develop a stakeholder engagement strategy

In the context of benchmarking, stakeholder engagement can be a strategic choice by policymakers, as well as good practice. Early and continual engagement with stakeholders ensures stakeholder acceptance and support for the exercise. It assists the administrator to better understand sectoral situations; make choices based on ground realities; pool industry knowledge; educate and inform participants; and mitigate possibilities of future disagreements. Many surveyed jurisdictions reflected on the utility of a concerted stakeholder engagement effort at the beginning of the benchmarking exercise.

A relevant authority will have to plan and execute a stakeholder engagement strategy to consult stakeholders on benchmark design (Step 1), aid implementation (Step 2-4), and facilitate further improvement to benchmarks over time (Step 5).

The experience of surveyed jurisdictions shows that that interaction and communication with stakeholders was a substantial challenge. Stakeholders need to be engaged to inform design choices —which sectors and what to benchmark. While studies can be commissioned to determine sectoral homogeneity, and perform disaggregation analysis, stakeholder consultation may be required to determine sectoral data feasibility, identify sufficiently similar comparable activities, and establish activity parameters for benchmarks. Californian policymakers found that it was important to engage with sectors to ensure they understood the process and the importance of the benchmarks. However, they also noted that it was a time-consuming process. In South Africa, although a default approach was proposed by the government, stakeholders had the liberty of proposing alternative approaches and the clay brick, cement, and sugar associations have taken on this initiative. Figure 12 outlines the key components of stakeholder engagement strategy, although policymakers should be mindful that individual approaches will vary depending on the specific context of each country and needs.

Figure 12: Components of a stakeholder engagement strategy

Why engage?	• What is the purpose of the stakeholder engagement? • What are the most important benchmarking issues that require stakeholder inputs?
Who should be engaged?	• How to define your priority stakeholders? • How to engage without affecting the decision making efficiency?
How to engage?	• What tools and instruments would be used for engagement? • What type of input is expected from stakeholders? • How will the stakeholder input be used for benchmarking?

Why engage?

A relevant authority must first reflect on the need for stakeholder engagement during each step of the benchmarking process and outline the main issues that require stakeholder inputs. Some common objectives include:

In Step 1: To support the decision of which sectors and what to benchmark by providing an understanding of what data is available, which activities are sufficiently similar for successful benchmarking, as well as an assessment of the feasibility of developing benchmarks and the appropriateness of the chosen activity parameters;

In Step 2: To provide data and information on existing data, to assess the quality of reported data;

In Step 3: To aid quality assurance of the analysis and encourage public participation;

In Step 4: To receive feedback on the calculated benchmarks; and

In Step 5: To encourage feedback on the impact of implementation of benchmarks and support benchmark review and updates

Who should be engaged?

A good entry point for defining priority stakeholders is to undertake a comprehensive mapping of all possible stakeholders who might contribute and be impacted by benchmarking. This mapping would include:

- ***Covered entities,*** for example, stakeholders directly affected by benchmarking;
- ***Stakeholders indirectly affected by benchmarking,*** for example, input-process-output chains of covered entities, public, civil society, industry lobby groups and associations;
- ***Experts*** from academia, industry, consultants, sectoral ministries; and
- ***Stakeholders relevant for implementation and outreach,*** for example, implementation agencies or media.

This initial long list of stakeholders can then be shortened for different steps through a ***prioritization exercise***. Prioritization would differ from country to country. The European Commission's minimum standards for consultation provide a useful framework for how countries might undertake such an exercise. This framework defines consulted stakeholders as those who: (a) are directly affected by the policy; (b) have a stated interest in the policy; (c) implement; and (d) have the relevant expertise on the subject matter. Further, stakeholders that represent public interest (civil society, citizens, etc.) are also increasingly included in stakeholder consultations. Care must be taken to balance the need for inclusiveness in the engagement with efficiency of decision-making and transaction costs involved.

Line ministries are important stakeholders in the benchmarking exercise. Bringing them on board early can be helpful in developing rigorous and relevant benchmarks. For instance, some ministries have established data collection systems that can be useful as a data source for the relevant authority. Even if such data is not of direct use for benchmarking, experts from line ministries can be requested to support cross-checking and validation of the information provided by entities. Their sectoral expertise can also be extremely relevant for benchmark review and update. For instance, South Africa's National Treasury, which is designing its carbon tax, has plans to engage with sectoral experts from the Department of Environmental Affairs, Department of Transport, and Department of Energy during benchmark development. This serves the purpose of cross-checking the collected data with existing records and will facilitate review of developed benchmarks. Central ministries such as the Ministry of Finance also are important stakeholders. For New Zealand's NGAs, the Treasury and Minister of Finance were engaged in the policy process; the Minister of Finance was also a joint signatory to the agreements.

How to engage?

Countries adopt a range of approaches for engaging with relevant stakeholders for benchmarking. Depending on the stage of benchmarking, different instruments can be used.

Targeted engagement approaches

These can take the form of *one-to-one meetings* with major players in the regulated sectors, industry associations, etc. Initial proposals by the relevant authority can also be shared with groups of participants in sectoral workshops and seminars. Targeted engagement is particularly useful in the planning stage of the benchmarking exercise (Step One), as this requires gathering information on sectoral contexts to design benchmarks. Most surveyed jurisdictions used this approach.

Questionnaires are another useful means of collecting information and gathering input, particularly in the data collection stage (Step Two). *Documented guidance/tools* on calculated benchmarks and their application for covered entities, such as Excel templates provided by the EU, New Zealand, and India, can be useful at the integration stage (Step Four) to understand the impact of benchmarking.

Finally, *technical working groups* involving relevant experts, as planned in Australia and South Africa, are important in the review, evaluation, and update of benchmarks (Step Five).

The relevant authority can provide a consultation document with details of the objectives and context of the consultation, itemized issues open for input, procedures for feedback and use of the engagement. In the case of technical working groups, clear terms of reference for external participants must be defined and the engaged experts should confirm there are no vested interests that could interfere with their judgement. Most targeted engagement approaches encompass input in the form of verbal comments, which are then catalogued in a consultation document or minutes of meetings.

Public consultations

Open consultations with the wider public may also be planned. These could be in-person public consultations to discuss proposals (e.g., initial coverage of benchmarked sectors in Step One) and information sessions to present outcomes (e.g., integration of benchmarks in Step Four). Broader open public consultations may serve to increase the political momentum prior to the launch of a scheme, and to receive input on any specific issues (e.g., review of benchmarks in Step Five).

Public consultations on benchmarking related aspects may be combined with other issues under the policy instrument. For example, in the EU ETS, public feedback was requested both for the legislative proposal on Phase IV revisions for specific technical rules including carbon leakage and including some suggestions for the revision of Phase III benchmarks.[20]

Online consultations

In addition to face-to-face consultations, online public consultations have become a common tool for reaching out to stakeholders for input or feedback. These can take the form of a questionnaire specifying the type of input required or simple requests for feedback without any formal structure.

Some key aspects must be kept in mind for planning a transparent online consultation. First, having an adequate timeframe for collecting responses is key to ensuring good participation. At the same time, the timeframe should not be so long that it interferes with timely decision making. Experience suggests that countries usually choose a four to eight-week response period. Australia provided a 30-day comment period for consultation on the draft guidelines for developing emissions intensity benchmarks, while the EU generally provides an eight-week period, extendable to more than eight weeks under certain conditions. In addition, a registration process with details of the stakeholder type is important for policymakers to understand the context of a respondent's input. For transparency, policymakers can consider adopting procedures to acknowledge receipt of responses and publish them in the public domain. Feedback on how their inputs were incorporated must be provided to stakeholders, either individually or in a common document. Online consultations can be particularly useful during the planning stage (Step One).

Overall, the critical element for stakeholder engagement is to balance comprehensiveness and transparency with efficiency in decision-making. Countries should follow their national legislative provisions for conducting stakeholder consultations (e.g., those established for environmental clearances) or otherwise establish specific procedures to ensure transparency. Box 7 outlines the European Commission's guiding principles and minimum standards for consultation.

[20] Further information can be found here: https://ec.europa.eu/clima/policies/ets/revision_pt

Box 7: EU guiding principles and minimum standards for stakeholder consultation

The guiding principles and minimum standards establish a framework for dialogue between administrative agencies and consulted stakeholders, and outlines requirements and practices that should be followed in consultations.

Four guiding principles are specified:

1. **Participation –** Inclusive participation of a large set of stakeholders.
2. **Openness and accountability –** Transparent approaches should be taken by the policymakers on stakeholder involvement and by consulted individuals/organizations with regards to the interests they represent.
3. **Effectiveness –** Need for early and continuous engagement with stakeholders and to follow the proportionality principle (i.e., the nature and depth of engagement must be proportionate to the policy impact).
4. **Coherence –** Consistency and transparency of consultation approaches taken, ensuring appropriate coordination and reporting.

All consultations must meet the following minimum standards:

- **Clear content of the consultation process –** This includes guidelines on the kind of information that should go into the advertising and consultation documents including outlining the context, objectives, timeframes and how the responses are taken into account in policy making, documentation of the issues for which inputs are sought, and next steps in policy development.
- **Consultation target groups –** Guidelines are provided for things to consider when defining the target group of stakeholders for consultations. These include the affected parties, implementation bodies, and parties that have direct interest in the policy. Additionally, consideration must be given to involving other participants based on impacts of the policy, specific technical knowledge, and experiences required for the asked questions, as well as wider representation of social and economic actors as appropriate.
- **Publication – The need for a 'single access point' is stressed. In case of EU, this is an online portal** *Your-Voice-in-Europe*. This can be supplemented with other means such as press releases, mailers, etc.
- **Time limits of participation –** As stated above, an eight-week period is considered standard for consultations (and 20 working days' notice for meetings) which can be extended in some conditions. The stress should be on good lead time for preparation with balancing effectiveness of decision making.
- **Acknowledgement and feedback –** The nature of acknowledgment depends on the total comments received, for larger feedback a collective acknowledgement could be sent. Feedback is usually in the form of a feedback document uploaded into the single access web-portal.

Source: European Union Commission, 2002

Table 11 presents an illustration of potential priority stakeholders and possible approaches to engagement during different steps in the benchmarking process.

Table 11: Stakeholders and Engagement

	Priority stakeholders	Engagement approaches
Step 1: Planning	• Representatives from regulated sectors • Sectoral associations • Consultants • Line ministries • General public and public interest groups	• One-on-one meetings with covered entities and their representatives • Expert groups • Technical studies • Public consultation • Online consultation
Step 2: Data Collection	• Representatives from regulated sectors • Industry experts • Consultants / survey agencies • Verifiers • Line ministries	• Questionnaires to gather data • In-person meetings • Outsourcing data collection • QA/QC

Step 3: Analysis	• Consultants • Line ministries	• Technical studies • QA/QC • Review of analysis
Step 4: Integration	• Industry participants and their representatives • General public and public interest groups • Implementing agencies, if different from relevant authority	• Public consultation • Online-consultation • Information sessions
Step 5: Monitoring and improvement	• Industry experts • Regulated entities • Line ministries • Consultants/academia • Auditors and Verifiers • General public and public interest groups	• Expert groups • Technical assessment • Auditing and reviewing monitored data • Reviews • Public consultation

3.2.3 Create institutional and legal capacity

Relevant authorities responsible for the design and implementation of the benchmarking exercise must have the institutional and legal capacity to perform this role.

Institutional capacity refers to the existence of institutions with the resources and mandate to carry out the benchmarking exercise. This includes the authority to mandate actions from private and public stakeholders involved in the benchmarking exercise. This may require "memorandums of understanding" between governmental departments, or contracts between government and entities, or legislative amendments to enforce the benchmarking process. Legal provisions which are relevant for benchmarking include rules for data collection and reporting from participants, the mandates and responsibilities of different sectoral agencies in implementation of the instrument, and the monitoring framework and scope for changes and improvement of benchmarks over time.

Legislative planning is a critical but time consuming component of the planning stage. The need for a new law or amendments to an existing one may be required. The level of detail covered, extent of stakeholder consultation, and time taken for legally enshrining the instrument and benchmarking provisions therein will differ across countries. Most often, the legal process is integrated into the law or policy for the instrument itself. Under the EU ETS, the entire benchmarking process and associated provisions are legally encoded as a Commission Decision on determining transitional union-wide rules for harmonized free allocation of emission allowances (EU 2011). In California, these fall under the Regulation for the California Cap on Greenhouse Gas Emissions and Market-Based Compliance Mechanisms.[21]

Including benchmarking in the legal framework of the instrument is beneficial for a number of reasons. This facilitates clear delineation of stakeholder roles, preventing subsequent conflict with respect to mandates and roles. It makes data collection easier, which is especially important when reporting is mandatory. Additionally, it outlines the consequences and penalties for non-compliance, laying a sound institutional framework for benchmarking. New Zealand's experiences with the Negotiated Greenhouse Agreements (NGAs) and ETS illustrate these advantages. The climate change legislation in 2002 laid the groundwork for the NGA program, however, the NGAs themselves were simply contracts between the government and entities. While lack of legislative support allowed the NGA program to develop quickly, it also made it more vulnerable to political change. This contrasts with the ETS, which was the subject of extensive and prescriptive legislation that set up the obligations and processes. It was passed

[21] Key regulations: (California, Cap-and-Trade Regulation section 95891(b), Appendix A: Additions and Amendments to Product-Based Benchmarks in the Cap-and-Trade Regulation, March 2014) (California, Cap-and-Trade Regulation section 95891(b), Appendix B: Development of Product Benchmarks for Allowance Allocation, July 2011) (California, Cap-and-Trade Regulation section 95891(b), Appendix C: New and Modified Product-Based Benchmarks, September 2013) For further information: https://www.arb.ca.gov/cc/capandtrade/capandtrade/unofficial_ct_030116.pdf

in 2008, amended in 2009 and 2012, and is currently under review, but its basic features have been kept in place.

Revisions may often be required as a scheme matures. In India's PAT scheme, the first phase highlighted necessary amendments to the Energy Conservation Act in order to implement the scheme, and the need for linking the verification processes under PAT with the Inspection Rules (2010).[22] In California, identifying and implementing amendments is part of the general public procedure. Stakeholders can make comments on any part of the Cap-and-Trade Regulation, and if policymakers determine that a modification is necessary, the change can be proposed when the regulation is opened for amendment.

[22] Based on literature provided by respondent.

STEP 1 Planning
STEP 2 Data Collection
STEP 3 Analysis
STEP 4 Integration
STEP 5 Monitoring and Improvement

4 Step Two: Data Collection

In the previous step, benchmarks are designed and capacity and resourcing plans are developed. In *Step Two: Data Collection*, requirements for data are specified and data collection approaches are chosen. The key activities and considerations in *Step Two: Data Collection* are presented below as an overview of the chapter.

Key Activities	Key considerations
Specify the data requirements ✓ Draws on the parameters chosen in the Planning step.	These requirements flow from design choices in Step One, namely the activity and impact parameters, and historical period.
Choose data collection approaches ✓ Three approaches can be distinguished according to whether they are based on existing or new data, and whether they are voluntary or mandatory.	Approaches have different implications for data relevance and resourcing. New data is generally more relevant than pre-existing data sets, and mandatory approaches increase the chance of obtaining sufficient representative data. Resource requirements are driven by the number of engagements with data providers. Experience shows that a combination of approaches is needed to address data availability issues.
Implement data collection approaches ✓ Prepare for data collection, assess IT requirements, data format, submission, and quality and verification requirements. ✓ Engage with the relevant data providers to request data (stakeholder engagement) and provide guidance.	Each approach has different requirements under these steps. Experience shows that careful planning is required for implementing data collection approaches assessing resource (human, financial) and time requirements, and addressing possible gaps. Treatment of sensitive or confidential data is also a key challenge, and the best practice is to agree an approach with stakeholders that addresses their concerns.

4.1 Specify data requirements

At this stage, policymakers will already have defined the data requirements broadly through deciding *what* and *wh*ich sectors to benchmark. These decisions will have resulted in the selection of benchmark *impact* and *activity* parameters. Impact parameters include GHG or CO_2 emissions, or indicators such as energy use. Activity parameters include outputs, indicated through production levels, services (floor area, kilometer traveled, etc.), or inputs (consumed fuel, heat, etc.).

The policymaker must now begin to further specify the data requirements for calculating the impact and activity parameters, and choosing the relevant historical period. The process of specifying data requirements is illustrated through examples from three surveyed jurisdictions whose experiences typify the process and cover a range of instruments. California, Australia, and India developed/plan to develop product-based benchmarks for three different instruments (ETS, Safeguard mechanism, and EETS respectively).

Specifying activity parameter data requirements

California, Australia, and India all chose/plan to choose levels of production of a particular product as the activity parameter. Determining overall production levels may require breaking down the activity into sub-activities. For example, in order to arrive at the total cement production equivalent to the major grade of cement in India, policymakers specified that data at entity level should cover these sub-activities: (1) total cement produced for each grade, (2) total clinker production, (3) details of additives used. Further specification is then required in terms of the scope and period of the activity data. For example, in California policymakers specified the data was to be provided at the entity level (i.e., covering the installations attributed to that entity) and that actual data had to be provided, rather than earlier forecasts or projections.

Specifying impact parameter data requirements

Impact parameters are determined by the instrument type. The California ETS and Australian Safeguard Mechanism are greenhouse gas-based instruments, and the impact parameter is therefore emissions unit of production. As an EETS, the Indian PAT focuses on energy efficiency, and the parameter is therefore energy consumption.

To assess the emissions per tonne of production, Australia proposes to refer to pre-existing data sets, and derived this from data reported under the National Greenhouse and Energy Reporting system. In California however, emissions had to be determined through calculation. To calculate emissions, policymakers required data on fuel consumption (for combustion emissions), process emissions, electricity generated and sold, steam purchased, and steam generated on-site and sold. In India, in order to arrive at energy consumption, policymakers required data on fuel consumption (by fuel type) and total electricity consumption and source (e.g., grid purchased or self-generated).

Specifying historical period

In California the default data collection period was 2008-2010 (a three-year average was taken), however this could be adjusted if these years did not represent normal operation for the entity. In India, data was also collected from 2008-2010 for the first cycle, but rolls forwards by one year for subsequent cycles. Finally, Australia proposes to select data from the three most recent years available (considered the most representative) when the instrument is implemented.

4.2 Choose a data collection approach

Once the policymaker has defined the data requirements, the next stage is to choose a data collection approach. To inform this decision, this section begins with an overview and comparison of possible approaches, and then guides the decision-making process. Note that further PMR guidance on these topics is available in the technical notes on data management (PMR 2013) and reporting systems (PMR 2016).

Overview of the data collection approaches

Three data collection approaches are considered here:

- Collection of pre-existing data;
- Voluntary collection of new data; and
- Mandatory collection of new data.

The main distinctions concern the type of data collected and the data provision obligation.

The data type can be pre-existing data sets or data specifically collected for the purpose of the benchmarking exercise (“new data”). Where pre-existing data sets are collected, the data providers are usually not the benchmarked entities themselves but, for example, private sector organizations or publically accessible databases. Where the data collected is new, the data providers are either the entities themselves, or intermediaries who represent these entities, such as industry associations.

The data provision obligation pertains to whether the data provision is voluntary or mandatory. Data provision is considered mandatory when an enforceable obligation is placed on the data providers. Where a mandatory data collection approach is used, the engagement with data providers usually falls within wider instrument compliance processes, otherwise it is based on bilateral engagements or studies commissioned by the relevant authority. Instruments that follow a voluntary participation model may also establish a mandatory data collection obligation for collecting data on the required variables. For this, a representative sample of entities in the sector is defined to collect data for the identified variables to establish baselines. In the past, some CDM standardized baselines[23] have used such surveys. Additional details on the three approaches are provided below.

Approach 1. Collection of pre-existing data sets

Data type	Pre-existing data, created for purposes *other than* the benchmarking exercise For example, verified GHG data, which is often collected as part of national GHG inventories, sector output data, or current standards on buildings energy performance. For adapted benchmarks, such data sets can be collected from other jurisdictions.
Data provider	Data-set owners include public sector stakeholders (e.g., the relevant authority in question or others) or private stakeholders (e.g., sector association) or publically accessible sources of data (e.g., online). This approach does not involve direct contact with the benchmarked entity by the relevant authority, as data is held by other providers.
Data provision obligation	Since data providers usually are not obliged to provide this data, it is termed voluntary. Note that although the collection of the data which formed the data set may be mandatory (e.g. GHG emission reporting for a national inventory), once collected, the transfer of this pre-existing data set to the relevant benchmarking authority could be voluntary, instead relying on an agreement between government departments. It is possible—but uncommon—that this data transfer be made mandatory.
Engagement with data providers	If this data is not owned by the relevant authority, the relevant authority is responsible for engaging bilaterally with the data provider.
Jurisdictional example	California used emissions data (some verified, some not) from the state's mandatory GHG reporting regulation (reporting began in 2009 for 2008 data) to help derive benchmarks.

Approach 2. Voluntary collection of new data

Data type	New data, mainly created for the purpose of the benchmarking exercise. For example, impact and activity parameter data collected for the selected historical period.
Data provider	The data providers are the benchmarked entities themselves, or intermediaries who represent these entities, such as sectoral or industry associations.
Data provision obligation	Since data providers usually are not mandated to provide this data, it is termed voluntary.
Engagement with data providers	The relevant authority is responsible for engaging directly with data providers, be these benchmarked entities or intermediaries (e.g., sectoral associations). An alternative is to outsource this engagement by commissioning studies or surveys to gather new data in order to calculate the benchmark.
Jurisdictional example	In Tunisia, data was requested from the cement sector on a voluntary basis, leveraging the good connections between the relevant authorities and sectoral participants. In the EU ETS, the methodology was defined by the policymakers but the sector associations themselves liaised directly with the companies in order to collect the data. Similar examples exist in Japan's JCM (S-CP).

Approach 3. Mandatory collection of new data

Data type	New data created for the purpose of the benchmarking exercise. For example, impact and activity parameter data collected for a selected historical period.
Data provider	The data providers are the benchmarked entities themselves or intermediaries who represent these entities, such as sectoral associations.
Data provision obligation	As data is provided through an enforceable obligation on participants, usually as part of instrument compliance process, it is termed mandatory.
Engagement with data providers	As the data collection approach is integrated into other instrument compliance processes, the relevant authority responsible for benchmarking may or may not be responsible for engaging directly with the data provider. Note that this is distinct from the MRV reporting requirements.
Jurisdictional example	In New Zealand, participants who wished to become eligible for free allocation were obliged to provide data to authorities. Similar examples exist in the California (ETS), India PAT (EETS) and UK CCAs (EETS).

[23] Please refer to the 'documents' submitted to the CDM executive board for the approved standardized baseline on 'energy use in the rice milling sector in Cambodia' as an example for the CDM approach. Accessible at: https://cdm.unfccc.int/methodologies/standard_base/2015/sb33.html

Comparing data collection approaches

This section compares the data collection approaches based on data relevance and resource requirements.

Data relevance

The chosen data approach has different implications for the relevance of the data collected. In this context, *relevance* is driven by whether the relevant benchmarking authority is specifying the data requirements, as they will able to make specifications that match their needs closely. In addition, a determination is made whether *sufficient representative* data is collected. The *sufficiency* will be determined by the response rate (i.e. responses received as a proportion of those requested). The *representativeness* is dependent on whether the sample of responses received covers a proportionally diverse number of the peer group. The three data collection approaches are reviewed below in relation to the relevance of the data.

Approach 1: Collection of pre-existing data sets

The pre-existing data set collected was created for purposes other than the benchmarking exercise, which may reduce the relevance of the data. For instance, the scope of coverage (geographic, sector, inclusion thresholds, timescales, and so on) may not be identical to that required, and may lead to gaps. In this case, the data should be complemented by other data sources. Regarding sufficiency, as the interaction with data providers is usually voluntary, the provision of the data cannot be assured since there is no obligation or enforcement for data provision. For instance, if the relevant benchmarking authority has to interact with another ministry, or another jurisdiction (e.g., in the case of adapted benchmarks) to obtain the data, this is usually done on a voluntary cooperative basis, as one ministry would not likely mandate the provision of data from another. Although this transfer could be made mandatory, examples are uncommon. Conversely, public data is easier to obtain.

Approach 2: Voluntary collection of new data

New data sets will be created for the purpose of the benchmarking exercise. While this may increase chances of data relevance, there may be issues with the representativeness of the data. This is due to the voluntary nature of the approach—relevant authorities may not be able to specify the survey sample. The sample may therefore be affected by self-selection bias (i.e., only those wishing to provide the information do so), and they may not be representative of the peer group. This will become clear in the data analysis stage, at which point additional data collection may be required to complete the sample. Regarding data sufficiency, a voluntary approach risks lower response rates because there is no obligation or enforcement of data provision. In Phase III of the EU ETS, data provision was not mandated and the response rate was not 100 percent in every sector. To address such gaps, the European Commission initiated studies to collect data from industry associations. However, experience in the EU shows that it would have been helpful if data provision had been mandatory, and this was implemented in Phase IV.

Approach 3: Mandatory collection of new data

New data sets will be created for the purpose of the benchmarking exercise. Further, a mandatory approach implies that the relevant authority is able to specify the sample size, which may in fact be 100 percent of the benchmarked entities. These two factors combined ensure that this is the most effective approach for ensuring data relevance. Moreover, a mandatory approach usually increases response rates, as there is an obligation and enforcement of data provision requirements. This increases the chances of data sufficiency. However, as the NZ ETS example demonstrates, there is still the risk of non-provision of data. This was the case for smaller entities, where response rates were less than 100 percent. Due to limited human resources and time, it was difficult to enforce data provision among smaller entities, and data quality enhancement had to be considered (see Step Three).

Resource requirements

The chosen data approach has different financial, technical, and human resource requirements. The level of resources required is mainly driven by the number of engagements required with data providers. Generally, the larger the number of engagements, the costlier and more time consuming the exercise.

Regarding human resource or personnel requirements, the relevant authority's[24] data collection team will require the following capabilities in order to carry out data collection: preparation of data collection templates, engagement and guidance of data providers, collection and manipulation of data, assessment of data. Regarding IT resources, an authority will require access to tools to allow the preparation of such templates (e.g., spreadsheets or similar), the engagement with stakeholders (e.g., email or online portals), collection (e.g., data submission software or platforms and data storage capacity) and data assessment (e.g., data manipulation software). The three approaches to data collection are reviewed below in relation to resource requirements and costs.

Approach 1: Collection of pre-existing data sets

The collection of a pre-existing data usually involves engagement with a limited number of data providers who are not the benchmarked entities themselves. Consequently, this approach is generally less costly in terms of personnel and IT resource requirements than the alternatives. Timescales will depend partly on the type of data provider. Where data is provided from within the same government department, or is publicly accessible, timescales are generally shorter than engaging with separate departments or jurisdictions. Regarding the most significant costs faced, surveyed jurisdictions agreed that personnel costs (internal staff or external consultant fees) were more important than technical / IT costs.

Approach 2: Voluntary collection of new data

The data providers are the benchmarked entities themselves, and collection may involve direct contact with benchmarked entity or could be outsourced to intermediaries. Since a greater number of engagements with data providers is required, personnel and IT resource requirements can be substantially greater than Approach 1. Outsourcing data collection to intermediaries or commissioning specific studies or surveys (e.g., sectoral associations, who engage with benchmarked entities on the authority's behalf) may be a cost-efficient alternative. Where data collection is not outsourced, the typically smaller sample size means this option is less costly than Approach 3, but more than Approach 1. This exercise can be quite time consuming (ranging from two to five months in surveyed jurisdictions). The experience of surveyed jurisdictions shows that timescales are a pressing challenge in this step, and there is a tendency to underestimate the required time. As above, personnel costs are the most emphasized by jurisdictions.

Approach 3: Mandatory collection of new data

As above, the data providers are the benchmarked entities themselves, and collection involves direct contact with the majority of benchmarked entities. The significant number of engagements with data providers means this is generally the most time consuming of the approaches. Regarding costs, while personnel costs will be high due to the number of engagements, if IT resources are integrated with other instrument systems, these may be limited. However, the integration of the data collection mechanism into the wider instrument compliance process requires careful planning from the very initial stages in order to minimize costs and delays. Surveyed jurisdictions (CA and NZ) identified timescales as a significant challenge under this approach, taking over six months. Most significant costs were personnel (requiring teams of four to five people in these jurisdictions, both internal staff and external consultants), rather than IT costs.

Table 12 summarizes the implications for data relevance and resource requirements for each approach. As described previously, data relevance is driven by two factors. First, if the authorities involved in the benchmarking exercise are able to directly specify their data needs this increases relevance. New data is generally more relevant than pre-existing data sets. The second factor is whether the data collected is sufficient and representative. Mandatory approaches increase the chance of obtaining sufficient representative data. Resources requirements—financial, technical, and human—are driven primarily by the number of engagements with data providers.

[24] As reminder, a distinction is made between policymakers and relevant authorities, both of which are involved in the benchmark development process. In this guide, a policymaker is a public-sector authority responsible for decision making. The relevant authority is responsible for executing the decisions made by the policymakers.

Table 12:Summary of data relevance and resource requirements under each approach

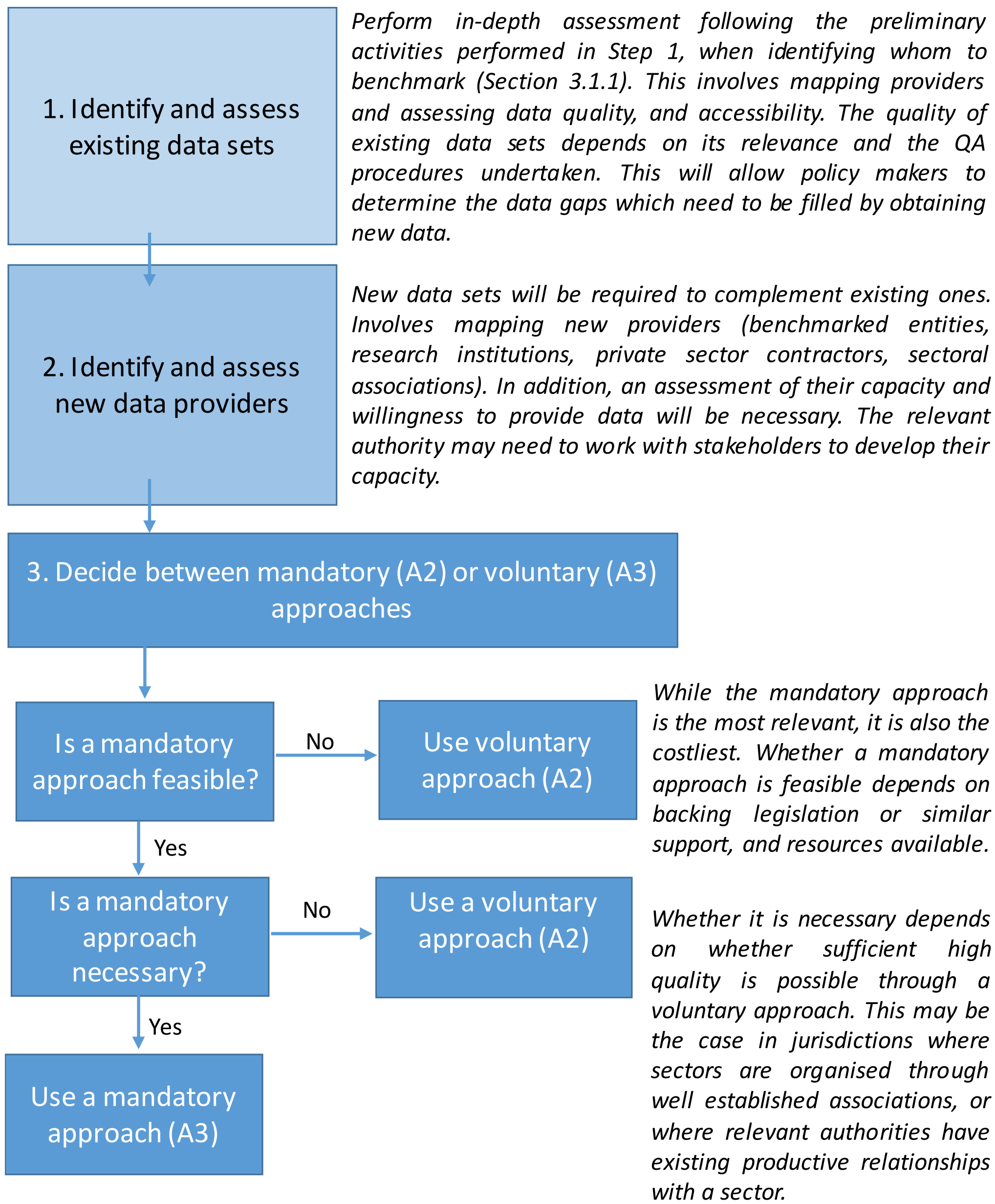
Approach	Data relevance	Resource requirements
Approach 1: Collection of pre-existing data sets	**Low**, due to the use of pre-existing data, and voluntary data provision.	**Low**, due to lower number of engagements with data providers
Approach 2: Voluntary collection of new data.	**Medium**. While the use of new data increases relevance, voluntary data provision may reduce response rates.	**Low/high**. Engagement with most benchmarked entities would lead to high costs, however if this can be intermediated, costs can be mitigated.
Approach 3: Mandatory collection of new data	**High**, due to the use of new data, and mandatory data provision.	**Medium/High**. Engagement with all benchmarked entities; may be mitigated by the integration of relevant costs (including IT) with other instrument systems.

STEP 1 Planning
STEP 2 Data Collection
STEP 3 Analysis
STEP 4 Integration
STEP 5 Monitoring and Improvement

How to choose the data collection approaches

The experience of surveyed jurisdictions shows that the most pressing challenge during this Data Collection step is limited data availability. To tackle this, the majority of jurisdictions combined more than one approach, specifically, the use of pre-existing data sets (Approach 1) with either Approach 2 or 3. Approach 2 and 3 are not usually used in combination, due to the voluntary or mandatory nature of the data provision requirement. Figure 13 provides a decision and activity tree to support the policymaker's choice of approaches. The outcome of this process is the selection of one or more approaches for data collection.

Figure 13: Choosing data collection approaches



4.3 Implement selected data collection approaches

Once the data collection approach or combination of approaches has been selected, it should be implemented by the relevant authority. Preparation for data collection involves an assessment of IT resource requirements together with specification of data formats, submission, and quality assurance requirements. Engagement with the relevant data providers will also be necessary to request data and to provide them with support during the data provision process. The experience of surveyed jurisdictions shows that resource limitations (human and technical skills) were pressing challenges for these activities. Careful resource planning, as described in Section 4.2, is suggested to avoid resource limitation problems.

Preparing for data collection

The relevant public authority must prepare for data collection through three main activities, which are detailed below.

1. Assessment of IT resource requirements

This involves an assessment of the suitability of the public authority's IT resources (see Section 4.2). In addition, an assessment of personnel capabilities and whether external support would be required (see Section 3.2.1). Wherever possible authorities should look to make use of the resources they have available for collecting, storing, and processing data.

2. Specification of data format and submission mechanism

The relevant authority must specify the type and format of data requested from the data providers, taking into consideration IT tools available to data providers. The format for collection involves specifying a data collection template and a submission mechanism. Examples of the data collection templates used in the UK's Climate Change Agreements, and New Zealand's ETS are provided in Annex A2. The UK's CCA template collects data on an entity's production levels and energy consumed. NZ's ETS template collects data on an entity's production levels, revenue and sales, and calculates emissions based on activity levels. Templates can be hard copies which can be posted, or rely on electronic submission via email or a designated data platform.

In Approach 1, since data format is pre-determined and engagement with providers usually voluntary, there is limited opportunity to specify data format and submission. Submission may be simpler if fewer providers are engaged. In Approaches 2 and 3, authorities will be able to specify the template for data and the mechanism for submission. For example, Tunisia prepared data capture tools for entities and New Zealand used an email based voluntary survey with a spreadsheet prepared by the authority, as highlighted in Annex A2. Depending on the volume of individual submissions, automated data submission platforms may be considered for efficiency. Approach 3 may present more opportunity for automation, as it is integrated into the instrument compliance process. For instance, in California data was collected through the existing mandatory reporting regulation (MRR), and verified by its third-party verifiers.

3. Specification of data verification requirements

The relevant authority may request that data providers perform quality assurance (QA) on the data before provision, for example, whether it must be third-party verified. In order to enhance data quality, the authority should provide clear guidance and support to data providers. For instance, if third-party verification is required, the authority will need to specify accredited verifiers and may need a verification standard to ensure quality and consistent verification.

Under Approach 1, there is limited or no opportunity to impose QA requirements from data providers, as the data has already been collected. This should be taken into consideration when deciding which data sets to use. Equally, as data provision is voluntary under Approach 2, authorities may not be able to impose QA requirements on data providers, although third-party verification may be required, as seen in the EU ETS Phase III. This approach was adopted to enhance the data quality, and therefore reduce the QA effort required by authorities. Under Approach 3, authorities typically require that providers arrange for third-party verification, and in some cases, this is complemented by verification by the authority's staff (California) or by a third party engaged by the authority (NZ).

STEP 1 Planning

STEP 2 Data Collection

STEP 3 Analysis

STEP 4 Integration

STEP 5 Monitoring and Improvement

Engagement with data providers

After preparation for data collection has been completed, the relevant public authority should engage with relevant data providers. This engagement involves two activities.

1. Requesting the data from the data providers

The complexity of this exercise may increase with the number of providers engaged and the complexity of the products / processes within a sector. Approach 1 usually involves engagement with a restricted number of voluntary data providers. These may be government stakeholders in the same or other jurisdictions or publically accessible data. Experience from jurisdictions indicates that bureaucratic red tape may be a challenge to data provision within public bodies.

While Approach 2 is also voluntary, it may involve engaging a large number of data providers (the benchmarked entities) as was the case in Tunisia, or a more restricted number of intermediaries, as was the case in the EU ETS. Encouraging voluntary provision may require leveraging good relationships with data providers, as was the case of the Tunisian authorities and their cement-sector stakeholders. In the EU ETS, the authorities liaised directly with sector associations via bilateral meetings, stakeholder events, expert groups, and email exchanges for Phase III. Formalized working groups and commissioning studies are further engagement options for consideration.

Finally, under Approach 3, a key concern is the communication of the mandatory data provision obligation. Industry consultations and other awareness raising and support activities can be conducted, as they were California. Californian industry associations and participants were consulted throughout, including methodology development. Although a time-consuming exercise, authorities recognized the importance of engaging with stakeholders in order to ensure understanding and gain their support.

2. Provision of guidance and support

It is best practice to provide guidance and support to data providers in order to maximize the quality and completeness of the data. In Approach 1, guidance is generally not required as data set has already been produced. However, it is very important for Approaches 2 and 3. Here, the creation of detailed guidance, and possibly helpdesks, is encouraged. Annex A2 provides examples of the guidance included in the data collection templates for New Zealand's ETS and UK's CCA.

Under Approach 2, a policymaker may choose to provide a helpdesk depending on the number of engagements with data providers and their level of readiness. Responsibility for this support may pass to intermediaries where these are contracted, as was the case of sector associations in the EU and South Africa. Under Approach 3, detailed guidance and helpdesks are usually provided for the compliance process. In addition, involving the data providers in consultations also increases awareness and aligns expectations, as was done in California.

Summary of key steps in implementation of data collection, under each approach is provided in Box 8.

Addressing data confidentiality concerns

The experience of surveyed jurisdictions shows that the treatment of sensitive / confidential commercial data in the public domain is a key challenge. It is best practice for the relevant authority to consult stakeholders and agree on a suitable approach to address such concerns. Under Approach 3, solutions to ensure data confidentiality include restricting access and viewership of the data to a neutral party such as the relevant authority, as adopted in California and New Zealand. In addition, while emission data is public in California, allocation data is not made public as output can be inferred.

In Approach 2, the commissioning of specific studies can provide a practical solution to data confidentiality concerns. Data can be handled by neutral intermediaries such as sector associations, and provided at a lower level of granularity or anonymized before further analysis and publication of the study. In addition, as in the EU ETS, a separation can be maintained between the list of installations and performance so the two could not be reconciled.

Box 8: Summary of key steps in implementation

Preparing for data collection

1. **Assessment of the authorities' IT resource requirements,** required under all approaches.
2. **Specification of data format and submission mechanism.** More detailed specifications are possible under Approaches 2 and 3 than under Approach 1; Approach 3 may also present more opportunity for automation of data submission.
3. **Specification of data verification requirements.** Under Approach 1 and 2 authorities are generally less able to specify such requirements, and undertake verification themselves. Under Approach 3, third-party verification is usually required.

Engagement with data providers

1. **Requesting the data from the data providers.** Approach 1 and 2 involves requesting data from voluntary providers, usually from a restricted number. In Approach 3, the provision of data is mandatory and usually integrated in system compliance mechanisms.
2. **Provision of guidance and support for data providers.** A very important step under Approach 2 and 3, where the creation of detailed guidance and helpdesks are encouraged. Generally, not required in Approach 1 as the data set is pre-existing.

5 Step Three: Data Analysis

In *Step Three: Data Analysis,* the quality and sufficiency of data collected will be assessed and improved if necessary before the determination of the benchmark value. Following this, ex-ante assessments of the benchmark can be performed prior to actual integration into the policy instrument. The key activities and considerations for policymakers in *Step Three: Data Analysis* are presented below.

Key Activities	Key considerations
Assess and improve data quality and sufficiency. ✓ Verifying data accuracy and relevance and assessing data gaps, determining whether there is enough data for a meaningful benchmark. **Determine the benchmark value** ✓ This is done using the chosen benchmark methodology, emission intensity curve, and the chosen stringency level. **Assess the benchmark** ✓ Perform ex-ante assessment of the benchmark. They can be assessed qualitatively, against the guiding principles, and quantitatively, modelling their costs and benefits.	Experience shows that resource limitations (human and technical) were pressing challenges and careful resource planning is required. Resources required depend on the volume of data and whether the data has been previously verified. Pragmatic approaches to reduce requirements exist, such as concentrating efforts on "leading practice records" where relevant. Where improvements are necessary but not possible, changes to benchmark design may be considered. Options include using alternative approaches, and existing benchmarks values may be considered (described in Step One).
Stakeholder engagement. ✓ Consult stakeholders on chosen methodology and data treatment, and for support on data quality and sufficiency assessment.	Stakeholder engagement is key at this stage. Further support may be required to assess and improve the data, and stakeholders will be interested in the method and outcome of the exercise. Finally, data confidentiality remains an issue, and authorities can address this by maintaining data analysis confidential or outsource the process to sector associations

5.1 Assess and improve data quality

This step begins with data collected in its raw form. In order to perform analysis on this data, a relevant authority will need a specialized technical team equipped with the right tools. The technical team will have to assess data quality (verification for accuracy and relevance) and determine whether any improvements are necessary. Generally speaking, tools for this analysis do not have to be very sophisticated. The majority of surveyed jurisdictions used/propose to use simple Excel-based spreadsheets as the main tool for the analysis (California, Japan, NZ ETS and NGAs, Tokyo, UK, Australia).

The level of effort a relevant authority will require for verification will depend on the volume of data (number of entities) and whether the data has been previously verified. For pre-existing data sets, it is necessary to check the quality assurance (QA) on the data. In the case of GHG emissions data, it may undergo QA by the relevant authority's technical team, in addition to requiring third-party verification. For new data collected, this will depend on whether the relevant authorities were able to request verification of the data, for instance by third parties, which may be possible under mandatory but not voluntary approaches. If data has not been verified for **accuracy**, it is the responsibility of the relevant authority to do so. Some key checks are required. Some examples include:

Plausibility checks. This involves comparing existing data with other data-sets and relevant sources of information. In the EU ETS, the voluntary approach to data collection for Phase III meant there was a possibility that some of the most efficient installations would not appear in the list. Checks were performed to ensure they were included, and a final cross check was performed against best available technology and literature.

Consistency checks. This refers to verifying that information reported by entities consistently used the correct units of measurement and baseline period. Further checks ensure consistency between data sets, especially where these were collected using different approaches. For example, checks ensure that the time basis for production data-sets matches that of emissions data-set, or are adjusted accordingly (e.g., calendar year or quarterly figures)

Anomalous data checks. This involves checking for outlying values of data, (e.g., too high or low). In California, staff reviewed for anomalous data and then contacted the facilities directly to correct errors or understand reasons for anomalies.

In addition, some authorities chose a pragmatic approach of paying special attention to data sets for the "**leading practice records.**" These are the records in the dataset for calculating the emissions intensity benchmark, that is, those with the lowest emissions intensity. Quality checks include:

- That facility boundaries are appropriate for calculating the benchmark and consistent within production and emission data-sets, and
- Data represents normal operating conditions.

Apart from checking for **accuracy**, it is also necessary to check for **relevance** of the data as required by the data specification in *Step Two: Data Collection*. An individual data record should correspond to emission intensity of producing one product, for a single benchmarked entity.

Since one benchmarked entity may produce several products, this step includes disaggregating data at multi-product facilities where more than one benchmark may apply to the same facility. This can be a time-consuming process. For example, in California, the authority's staff had to engage bilaterally with each multi-product stakeholder to carefully define the boundaries of the production process under each distinct benchmark.

In addition, the following checks ensure consistency with specifications:

- **Scope.** In cases where existing GHG emission data sets are being used, they must be consistent with the scope of emissions defined within the benchmark boundary, and in some cases adjustments may be required.
- **Historical time period** or the baseline period to which the data corresponds.
- **Units of measurement** are known, consistent and meet the measurability requirement for production variables, or can be adjusted so that its units do meet this requirement.

After the assessment has been performed, if there is any data which is of unacceptable quality, further efforts may be needed to improve data quality, such as: following up with data providers for further clarification; making the adjustments for scope, units, and measurement period; or performing further data collection activities.

5.2 Assess and improve data sufficiency

Once data is determined to be of acceptable quality, the sufficiency of the data must be assessed. A relevant benchmark can only be derived if it is based on sufficiently representative data. Assessing data sufficiency involves questioning whether there is enough data in order to derive a meaningful benchmark.

Data may be insufficient in two circumstances. Firstly, where there are data gaps, for example, if the benchmark is to be based on the best 25 percent of peers within a sector, but the data set only covers 90 percent. In this case, a pragmatic approach is to focus on collecting sufficient data for the most important facilities—those representing best practice on which the benchmark will be derived. Australia, proposes to check the accuracy of the data collected from "leading practice" (top-performing) data records, as emission intensities from these facilities were more likely to form the benchmark value. Otherwise, further data collection may be required.

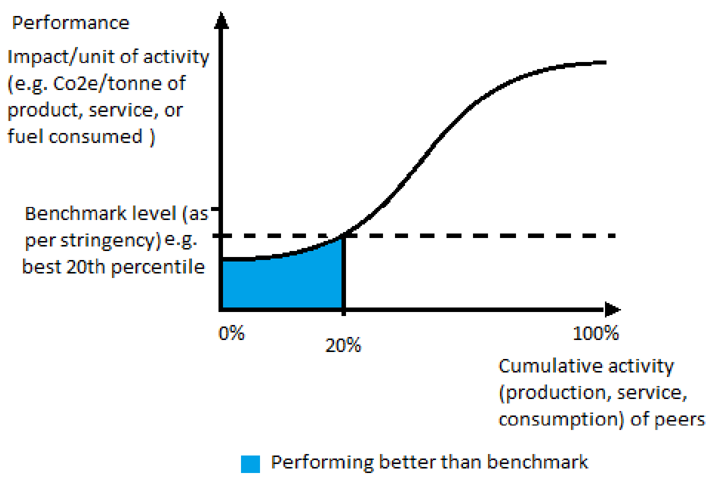
Secondly, in sectors where there are too few facilities, data sets may not be large enough to determine common or best practice. Three surveyed jurisdictions had a minimum requirement of two facilities per sector in order to determine a benchmark locally. Where the number of facilities is low, data sufficiency may be improved by using a longer baseline period as this will increase the data points in the set. Alternatively, it is possible to use international best practice data to set the local benchmark (similar to best available technology approach), as was done in NZ.

5.3 Determine the benchmark value

Once the data quality and sufficiency have been checked, the next step is to determine the benchmark value. This involves calculating the emission intensities of benchmarked entities according to the methodology and stringency levels chosen in *Step One: Planning.*

Figure 14 illustrates the calculation of an output-based benchmark through the use of an emission intensity curve. Here the activity parameter is production level and the impact parameter is CO_{2e} emissions. At this stage, data on emission intensity of production (CO_{2e}/tonne product) for all peers within the benchmark exercise will be aggregated and plotted on an emission intensity curve, in ascending order of intensity. The next step is to refer to the selected benchmark stringency level (e.g., top 20th percentile), and the corresponding benchmark emission intensity is found. The benchmark value will represent the emissions intensity of the top 20th percentile of peers carrying out a particular activity in a particular historical baseline period.

Figure 14: Example of emissions intensity curve and calculation of benchmark value



Source: Author's illustration

Since this benchmark is based on historical data, it can be termed an ex-post benchmark. The underlying assumption in using such ex-post benchmarks is that the past environmental performance is representative of current and future performance. This implies that the benchmark will be used in an unmodified state, irrespective of whether it is representative of the future. In theory, this data can also be used to forecast benchmark intensity for a future period, making adjustments for instance for technological progress, and how this may change the environmental performance per unit of production. Adjustment factors include forecast improvements in carbon efficiencies of technologies, and may be obtained from sectoral roadmaps or expert assessments of sector technological change. This represents an ex-ante benchmark; however, it is not a common usage of benchmark analysis.

5.4 Assessing the benchmark

At this stage, an ex-ante assessment of the benchmark can be performed to check if the benchmark is in line with the principles defined in Section 2. The four principles are alignment with policy objectives, robustness, fairness and conservativeness. A qualitative assessment framework for assessing the benchmark is presented below.

Is it in line with policy objectives?	✓ Are the impact parameters in line with the climate policy instrument? ✓ Are the activity parameters in line with the desired scope of activities? ✓ Does the stringency level chosen reflect policy objectives?
Is the benchmark robust? Function of measurability, transparency, and relevance of the benchmark.	**Accuracy** ✓ Has the data been verified, and does it conform faithfully to requirements? ✓ Have plausibility, consistency, and anomaly checks been performed? **Measurability**. ✓ Is the benchmark based on objective parameters, such as on quantitative, physical metrics? ✓ Is the data of high quality, for instance, verified by third parties? ✓ Is the data analysis process robust, ensuring data quality and sufficiency? **Transparency**. ✓ Are the calculation methods and benchmarks values available for public scrutiny? **Relevance**. ✓ Are the chosen benchmark parameters representative of environmental performance? ✓ Is sufficient data collected from a representative sample of the peer group? ✓ Is the benchmark expected to remain relevant as time passes, and are necessary improvements and reviews planned?
Is the benchmark fair?	✓ Is performance being compared across the correct set of peers – i.e. actors which are sufficiently comparable with respect to the parameters of the benchmark? ✓ Have efforts been made to use the same benchmark methodology, as far as is justifiable? ✓ Have efforts been made to define benchmarks so they cover as many peers as possible? ✓ Have all peers been treated equally in the data collection and analysis stage?
Is the benchmark effective?	✓ Have output benchmarks been preferred (where feasible) to alternative approaches such as fuel, adapted, and heat benchmarks? ✓ Have the majority of emissions intensive processes been covered by output benchmarks? ✓ Has the differentiation of product benchmarks been limited while treating entities fairly?
Is the benchmark feasible?	✓ Has a conservative approach to balancing the principles with practical considerations, of limited resources and data, been taken?

This qualitative assessment can be complemented by additional quantitative assessments of the cost-benefit impact of applying these benchmarks, or similar socio-economic impact assessments. In Japan's JCM, such a study was commissioned to analyze the impact of benchmarks on net emission reductions, requiring additional analysis and tools, such as the modelling of future activity and emission levels. However, this lies outside the remit of calculating the benchmark.

5.5 Stakeholder engagement

Stakeholder engagement is a key issue throughout *Step Three: Data Analysis.* In some cases, stakeholders are engaged to support the data quality and data sufficiency assessment and for improvement of the process. In addition, there is generally significant interest in the methodology used and the treatment of the data. This includes the level of public scrutiny the data will be subjected to, mainly for reasons of data confidentiality. Box 9 provides examples of the types of engagement used to assess and improve data analysis.

Box 9: Stakeholder involvement in assessment and improvement of data

- In California, stakeholders checked the data to ensure that the regulator was doing the calculations accurately.
- Under the NZ NGAs, stakeholders reviewed and agreed on the analysis. However, in the case of disagreement, third-party verification was performed.
- In many methodologies developed under Japan's Joint Crediting Mechanism, stakeholders shared their views through a follow-up survey.
- UK's CCA targets were set by negotiation between government and sectors, but subject to third-party review.
- Australia intends a review process which will consult stakeholders.

The level of public scrutiny is linked to the issue of confidentiality of the data. Stakeholders are naturally concerned about the level of public scrutiny that sensitive commercial data will receive, particularly production data. Jurisdictions have dealt with the issue of ensuring data confidentiality in a variety of ways. The approach taken in California was for the responsible authorities to perform the analysis (which may involve the use of contracted consultants) themselves and only allow access to raw data and analysis to the concerned stakeholders. The data may then be presented for public scrutiny at a low level of granularity. An alternative approach is to outsource the analysis to sector associations so that data is not transferred, and to publish data only at a low level of granularity. Examples include the EU ETS and South Africa, where guidance was produced by the relevant authorities, and the sector associations themselves developed / are developing approaches to gathering and handling the data.

Overall, for the sake of transparency, public access to the data and outcomes of the analysis are encouraged, and covered more fully in *Step Four: Integration*.

6 Step Four: Integration

Integration is the final benchmarking step before scheme implementation. Here the relevant authority uses the benchmarks developed in the previous step to determine system targets and thresholds, and apply them to determine the level of distribution of system benefits and obligations. In order to determine distribution levels, additional data collection and calibration of distribution levels may be required.

The key activities and considerations for policymakers in *Step Four: Integration* are presented below.

Key Activities	Key Considerations
Apply benchmarks in the policy instrument to determine system benefits and obligations.	
✓ Arrange additional data collection for benchmark application, where necessary. There is an ongoing requirement to collect impact and activity data to determine distribution levels.	Activity parameters used in this step should correspond to those defined in Step 1. Data collection process for activity parameters is at the level of facilities and additional to that collected for benchmark development in Step 2 (e.g., different time periods, quality of data). Yearly data collection can be integrated into the reporting requirements of the scheme. Decisions to this effect must be made in the beginning of the benchmarking exercise or during the instrument's MRV design.
✓ Calibrate distribution levels, where needed. Technical and additional factors may be needed to calibrate distribution levels.	Adjustments are made only to the final distribution levels. No changes are made to the benchmark. Decisions on needed adjustments should be taken in the beginning of the benchmarking exercise or even before, during the design of the climate policy instrument.
Stakeholder engagement ✓ Communicate benchmarks. ✓ Provide guidance on application. ✓ Address stakeholder grievances.	Experience gained from jurisdictions highlights that early and continual engagement with stakeholders in the initial steps can reduce the need for outreach during integration.

6.1 Applying the benchmark in the policy instrument

Once benchmark values have been calculated following steps 1-3, the policymaker will use these along with data on activity parameters to determine the distribution levels for benefits to or obligations on the entities participating in the instrument. Benchmarks are applied for different purposes in climate policy instruments. As discussed in the beginning, these are:

Determining ETS allocations. Where benchmarks are used to determine distribution levels, the benchmarks are applied together with activity parameters data to determine the level of allowance distribution.

Determining carbon tax thresholds. For CT, benchmarks may be applied to determine eligibility for a rebate, or the proportion of tax rebate allocated. The entity liable for the tax would need to determine its performance against the benchmark during future tax years, involving verified emissions and activity parameter data.

Determining EETS system targets or S-CP crediting thresholds. Benchmarks can be used to set thresholds or define targets for an environmental performance parameter that an entity must meet (e.g., emissions in an S-CP and energy use in an EETS). In doing so, the environmental performance of the entity is compared to the benchmark. For S-CPs, the entity receives emission reduction credits if it performs better than the benchmarked crediting threshold. In an EETS, penalties may be set for those who don't achieve the target energy performance. The activity parameters used alongside benchmarks can be outputs, indicated through production levels and services (e.g., floor area, kilometer traveled etc.) or inputs (e.g., consumed fuel, heat etc.). These should correspond with what was chosen during Step 1 to define the benchmark.

Arrange additional data collection for benchmark application

For defining distribution levels, data must be collected from the covered entities on their activities and the resulting impacts on an ongoing basis. In this regard, ETSs are notable and discussed further. In ETSs, activity data is used for determining annual allocations. In principle, two allocation approaches can be applied. Both involve gathering actual activity data, but for different timeframes. These are:

- Using historic activity data to determine fixed ex-ante allocations. Each installation would receive an allocation that would be derived from its historic activity multiplied by the benchmark. Allocations would be known in advance of each emission year and be fixed.
- Using actual activity levels to define ex-post allocations. This could either involve a full ex-post allocation or an ex-post adjustment to a provisional allocation set ex-ante. In this case, the level of allocation at installation level and the cap for the system would be uncertain at the start of each emission year.

For a system employing ex-ante allocations, policymakers may use the activity data collected for defining benchmarks in Step 2 for the first allocation. For successive allocations, new data needs to be gathered on a regular basis. This data collection can be integrated in the MRV compliance mechanism of the instrument and decided in the beginning of the benchmarking exercise or during the instrument's MRV design. Jurisdictions' experiences highlight some common situations when a separate activity data collection may become necessary for the first allocation as well. These are:

- Activity data for calculating allocations might be needed for a longer historical period than that used for benchmark determination. For example, the EU-ETS used four-year activity parameter data from facilities (2005-08), while benchmarking was done using two-year data (2007-08). This is because unusual operations (seasonal/annual fluctuations) at individual installations have a small effect on the benchmark value as it needs to be representative of a population of installations. However, representativeness of historic activity parameter data from facilities becomes critical for accurate determination of allowances to be allocated to the facility. Hence, longer historic periods may be used, or rules set to exclude unrepresentative years (e.g., when the installation may be operating at low or zero throughput) for facility level data collection.
- Data collected during benchmark development may not cover all installations, or may be based on existing benchmark values (from other jurisdictions or existing best available technology literature), or may have been provided in an anonymized manner. In these instances, a policymaker would need to carry out additional data collection for the first allocation as well.

Finally, additional data collection arrangements may be needed when benchmarks are applied to new or modified facilities for which no historical baseline exists. This will depend on the allocation method for new entrants. Taking the industrial sector as an example, one may utilize planned or intended production *capacity* levels and assumptions about *utilization* of that capacity to derive an activity level, and hence an allocation (e.g., in EU ETS). Alternatively, anticipated activity may be used for a provisional allocation, which is then subject to an ex post adjustment once actual activity levels for the year in question are known (e.g., in New Zealand).

Calibrate distribution levels

Once the final distribution level has been calculated, relevant authorities may require some adjustments to these values (see Table 13). In some cases, distribution levels are set differently to address the inherent issues with respect to the assessment method used for defining distribution levels. In others, adjustments are done to accommodate external factors or implement specific policy goals.

To address issues inherent to the methods used, relevant authorities often define **technical factors** to calibrate the final level of distribution of system benefits or obligations. In an ETS, technical adjustments are needed when there is a difference between the overall level of free allocation for a sector based on the emission cap and the aggregated allocations calculated based on facility level data (i.e., benchmarks multiplied by activity level). For instance, the EU ETS uses a cross-sectoral correction factor to reduce the number of allowances in case the bottom-up allocation exceeds the top-down allocation limits defined for the system. The factor also ensures that the overall cap is not exceeded. The ex-ante allocation approach (with technical adjustment factors) is commonly adopted in jurisdictions. However, adjustment can also be done to maintain a sectoral cap in a dynamic way, as discussed in the ex-post allocation approach followed in the IFIEC method (Schyns 2006 in Wesselink et al. 2008, pp. 11–14). In this method, first a sectoral cap is set for a defined compliance period. Next, based on activity data, scenarios for annual emission caps and yearly benchmarks in that period are estimated. For a particular year, allocations are distributed (tentatively) using historical or estimated activity data and that year's benchmark. In the next year, the estimated allocations are checked against actual production figures. If activity levels are different from estimated values, these are subtracted or added (in even parts) from the cap of the remaining years. The respective benchmarks are also adjusted accordingly. Therefore, the overall cap is maintained in a dynamic manner.

In an EETS, technical calibration may be needed to minimize the impact of external factors on the target. For instance, a facility may get undue advantage or disadvantage in achieving its target if differences exist between its base year and target year operating conditions. In such situations, relevant authorities use normalization factors to standardize the operating parameters in the target year with respect to the base year (e.g., in India's PAT scheme). In the PAT scheme, system targets are based on specific energy consumption (SEC) of facilities. For the target year, the SEC calculation is normalized to nullify the effect of external factors on performance of the entity. These factors include changes in the product mix, capacity utilization changes, changes in fuel quality, import/export of electricity etc. (MOP 2015).

Policymakers may also apply some **additional adjustments** to the final allocation / rebate / credits to accommodate specific policy objectives. Some examples of the policy goals fulfilled through additional factors are discussed below:

Addressing carbon leakage. Carbon leakage is the risk of increase in total emissions of entities who may move their businesses out of jurisdictions with a stringent (climate) policy.[25] A carbon pricing policy like ETS impacts the cost of production of the covered entity, which forms a key determinant of competitiveness for carbon intensive firms. Further, the lower the ability of a firm to pass on the carbon pricing costs without significant loss in market share, the higher the risk of leakage. Thus, carbon intensity and trade exposure of a sector generally define its vulnerability to carbon leakage (PMR 2015a). In addressing carbon leakage, the calculated benchmark value itself is not adjusted. Instead adjustments to the calculated allocations are made to give more allowances to sectors vulnerable to carbon leakage. For example, in California ETS, allowances are adjusted using an 'industry assistance factor.' This factor is derived from a sector's emissions leakage classification (high, medium, or low risk) based on emission intensity and trade exposure. The level of the factor is decided by the regulator. For the initial compliance periods, all risk categories received 100 percent free allocations in California. For the third compliance period, medium and low risk categories are expected to receive lesser free allocations than high risk categories. Similar relaxations are provided in the EU ETS where sectors at carbon leakage risk receive 100 percent of the total allocations—calculated using the benchmark—for free, while for other sectors free allocation is below 100 percent, reducing from 80 percent in 2013 to 30 percent in 2020 in Phase 3. (EC, 2016a).

Adjusting for new entrants, closures, and changes. As mentioned earlier, benchmarks can be used to determine allocations or allowances for new participants in an instrument and for changes to existing facilities (including closure or reduction in activity). These adjustments could be made annually or on an ad-hoc basis, such as at the point in time when facilities change their operations. The rules for making such adjustments may include adjustment factors. For example, within the EU ETS sub-installations with activity reductions of between 50 to 75 percent, 75 to 90 percent or >90 percent of their initial activity levels (i.e., allocation baseline) receive 50 percent, 25 percent, or none, respectively, of their initially calculated allocations each year.

Once all relevant factors are applied, the final allocation/ tax rebate/ credits can be calculated. Taking the example of ETS again, a calibrated distribution level (i.e., an allocation) is illustrated in Box 10. Some examples of calibrations done by different jurisdictions are outlined in Table 13.

---

[25] A detailed discussion of carbon leakage can be found in PMR, 2015; 'Carbon Leakage: Theory, Evidence and Policy Design.' *Technical Note 11,* World Bank Partnership for Market Readiness.

Box 10: Application of benchmarks in Emission Trading Schemes

In an ETS, benchmarks are used to determine the level of allowances to be distributed. The level of allowances allocated is calculated using facility level data of the chosen activity parameter, and benchmark.

allocated allowances = benchmark * facility activity parameter * factors

Where,
allocated allowances = Emission allowances given out for free to a facility (e.g. in t CO_2 / year)
benchmark = Benchmark for the activity indicator (e.g. t CO_2 / t product)
facility activity parameter = Activity parameters may be expressed in units of output produced, or inputs consumed. (e.g. t product / year, measure of service / year)
factors = Technical factors or additional adjustments to accommodate method related issues, effect of external variables or policy goals

Allowance allocation is carried out for all participating facilities on an annual basis.

Table 13: Examples of jurisdictions using adjustments

Adjustments	Examples of jurisdictions implementing the adjustment
Cross-sectoral correction factors	EU ETS
Normalization factors	India's PAT scheme
Adjustments to address carbon leakage	California ETS, EU ETS
Adjustments to reward good performers	South 'Africa's Carbon Tax
Adjustments for new entrants, closures and changes to operations	EU ETS

6.2 Stakeholder engagement

In this step stakeholder engagement involves communicating benchmarks, providing guidance to familiarize stakeholders with the application of benchmarks, and addressing stakeholder grievances.

Communicate benchmarks

A key aspect of the integration step is to effectively communicate the benchmarking values to covered entities. Jurisdiction experiences differ on the extent of engagement, and approaches and measures taken to gain acceptance of benchmarks.

Provide guidance on applications of benchmarks

Some jurisdictions develop ***guidance and tools*** for stakeholders to acquaint themselves with the benchmarking values and their usage. Jurisdictions implementing ETSs have developed tools for covered entities to calculate their allowance allocations. For example, the European Commission has provided an Excel template application for incumbents and new entrant allocations for relevant data collection as well as extensive guidance on the specific rules that apply (EC 2016b). The tool includes calculation of annual allowances allocated freely. New Zealand ETS provided excel based spreadsheet templates to participants for determining their eligibility and allocation baselines, but allowed them to apply for annual allocations online (Ministry of Environment 2016). In project based crediting, many national agencies publish and update the emission factors for grid based electricity generation. The UNFCCC secretariat has also developed guidelines for development of standardized baselines for similar mitigation activities (project types) in 2011.

Apart from providing guidance, policymakers may need to undertake ***direct consultations*** with covered entities to develop their understanding of the benchmarks and the overall metric. These could be face-to-face discussions in bilateral or group meetings and information sessions.

Address stakeholder grievances

Stakeholder engagement may also be needed to address pending grievances. However, the level of effort needed towards this depends upon the nature of stakeholder engagement in previous steps. Some jurisdictions (e.g., NZ and CA) had limited need to conduct an elaborate stakeholder engagement in this step as most efforts were done during benchmark development when industry specific information and data were needed from stakeholders. As private-sector grievances were already addressed in previous steps, there were fewer disagreements in the integration step. Further, some jurisdictions had already included benchmarking in the instrument's regulations before reaching this step, which further reduced any disputes.

Transparency in benchmark development process, adequate engagement with regulated entities from the beginning, and embedding benchmarking in the instrument's legal framework can decrease the costs and efforts towards developing consensus during the integration step.

7 Step Five: Monitoring and Improvement

Once the benchmarking exercise is implemented, regular reviews and update of benchmarks become essential to ensure their continued relevance, stringency, and fairness. This is because benchmarks generally use data from a representative historical baseline period to reflect the sectoral characteristics of that time. Yet the sector performance will change, for example, as efficiency improvements are made. *Step Five: Monitoring and Improvement* involves design decisions on the benchmark update approach, the development of a monitoring and review plan, the engagement with stakeholders regarding the plan, and potential updates to the benchmarks.

The key activities and considerations for policymakers in *Step Five: Monitoring and Improvement* are presented below.

Key Activities	Key Considerations
Design the benchmark update approach	
✓ Choose suitable approach for updating dynamic benchmarks. ✓ Benchmark update frequencies can be predefined (ex-ante update) or determined through ex-post reviews. ✓ Decide what circumstances will trigger benchmark updates.	Dynamic benchmarking allows for continual assessment of improvements. Ex-ante update approaches aim to push sectors to improve faster by pre-defining improvement timeframes.
Develop a monitoring and review plan	
✓ Determine the frequency of review of benchmark values. ✓ Plan data collection for monitoring based on approaches established in Step 2. This involves defining data requirements and data collection approaches. ✓ Plan for data review as per approaches established in Step 3 for analyzing data quality, and recalculate the benchmark. ✓ Decide what circumstances will trigger benchmark updates.	It is pragmatic to integrate the timeline of benchmark review and required data collection for update into the wider compliance and MRV process of the policy instrument where possible, as this saves resources. Not all reviews will lead to a benchmark update. In some cases, only minor adjustments might be required.
Stakeholder engagement	
✓ Communicate with stakeholders and provide guidance on monitoring and review procedure. ✓ Consult stakeholders in the review and update process.	Stakeholder engagement at this stage is as important as in other steps to both guide them on the monitoring plan, and consult them on the benchmark update.

7.1 Design the benchmark update approach

A benchmark is a useful metric as long as it is representative of its sector. It needs to be recalculated if unrepresentative, else there is a risk of compromising the integrity of the policy instrument. Therefore, benchmarks should be updated regularly during policy implementation. This approach is called **dynamic benchmarking**, as opposed to calculating the benchmarks once and not changing them thereafter (**fixed benchmarking**). The dynamic approach ensures that the benchmarks in use are relevant for the sector, strict enough to lead to real environmental impact, and reflect the same ambition level for all covered sectors (i.e., are fair). A dynamic benchmarking approach sets a performance improvement trajectory for the participating sectors. Deciding the benchmark update approach early on and communicating to the participating sectors is critical as it gives a clear policy signal to participating businesses.

Jurisdictions may decide to adopt a dynamic approach in the planning stage itself and incorporate it in the legal framework for the instrument. Regular reviews of benchmark performance are a necessary precondition to check for relevance of benchmarks. Frequency of update and the rate of change of the benchmark in each update can be pre-fixed (i.e., ex-ante updates). Alternatively, benchmark change can be defined after reviewing relevance of existing benchmarks during the scheme's implementation without prescribing any changes in advance (i.e., update based on ex-post reviews).

- The **ex-post updates** are done based on results of a review of existing benchmarks. Jurisdictions take varied approaches in planning such reviews and on whether a review will lead to benchmark update. For instance, in the UK's CCA scheme, sector commitment targets (i.e., agreed-upon percentage improvements) were reviewed in 2004 and 2008 and adjusted, depending on past performance, to ensure they remain realistic but challenging.
- In **ex-ante updates**, the frequency by which the benchmark value would change is pre-decided. It may also specify conditions under which changes to the benchmarks could be made. Implemented examples are scarce, however, this approach has been proposed by the EU for the fourth phase of EU ETS.

As with ETS, benchmark updates for setting crediting thresholds can be either a fixed or dynamic process. Here, benchmark revisions can be linked to crediting periods or decided between funders and host countries.

Decide what circumstances will trigger benchmark updates

Whichever approach is taken to benchmark updates; policymakers must consider up-front the key conditions under which a benchmark would be changed. This helps to inform the benchmark data monitoring plan and stakeholder communication strategy. The following are examples of changes that could trigger a benchmark update.

Changes in the policy objectives. Under the bottom-up commitment regime set in the Paris Climate Agreement.[26] benchmark updates for carbon pricing instruments can be one of the vehicles for ratcheting-up the ambition in an instrument.

Changes in data underlying the benchmark values. This can include issues such as:

- Errors in previous calculations of benchmarks (e.g., proposed in Australia);
- When better quality data for production variables becomes available than was used during benchmark calculation. This is specifically relevant when the benchmarks are developed using less than optimal data (e.g., when the number of facilities used in benchmark development were lesser) or when proxy approaches were used in the absence of data (e.g., the reserve approach[27] planned in the Australia);
- When stakeholder feedback reveals the need for changes. An experience of this comes from California's ETS, where stakeholder feedback was extremely useful in modifying the benchmarks originally proposed for all covered sectors;
- When there are changes in international standards such as the Global Warming Potentials of gases used in benchmarks calculation (e.g., proposed in Australia).

Changes in sectoral emission efficiencies must also be monitored at regular intervals to check if the sectoral context has changed significantly. Improvements in carbon efficiencies can be assessed from sectoral roadmaps that project future improvements. The ex-ante update rates mentioned in the EU legislative proposal are derived from historical annual improvement rates in the period 1990–2010. An expert assessment of sectoral technological change might be an option if such detailed assessments are not available.[28]

Further, a policymaker can define certain ***thresholds that trigger an update***. For instance, in Australia, a benchmark is updated if the recalculated value decreases or increases by five percent of the current

[26] By the beginning of December 2016.

[27] The Australian scheme proposes to use a 'reserve approach' to benchmark calculation when the available production data is insufficient. The reserve approach is flexible with the aim to develop a benchmark indicative of what a benchmark would look like.

[28] Example: EU commission's proposal an annual flat rate reduction of benchmark values is based on technological progress achieved in a sector. This ex-ante rate will be cross-checked with sectoral data and based on the actual situation of a sector, three categories are proposed. Source: https://ec.europa.eu/clima/policies/ets/revision/documentation_en.htm

STEP 1 Planning
STEP 2 Data Collection
STEP 3 Analysis
STEP 4 Integration
STEP 5 Monitoring and Improvement

value, and when either the benchmarks do not reflect current practice (due to erroneous calculation) or when a reserve approach was taken for calculating current benchmarks (Australian Department of Environment 2016). Thresholds can also be set for production variables (e.g. percent change in production levels) that feed into the benchmark calculations.

7.2 Develop a monitoring and review plan

Having decided the benchmark update approach, a monitoring and review plan can be developed. This will determine the detailed data collection and review processes.

Data requirements will be as per those specified in Step 2, and they will follow from the impact and activity parameters of the benchmarks. Impact parameters include GHG (or CO_2) emissions, or indicators such as energy use. Activity parameters include outputs, indicated through production levels, services (floor area, kilometer traveled, etc.), or inputs (consumed fuel, heat, etc.).

Relevant authorities should remember the following when planning the monitoring exercise for benchmark review and update:

- Outline the key variables for monitoring based on variables identified in Step 2.
- Understand the state of current data reporting under the scheme and elsewhere.
- If the required data is already reported under the scheme, check its completeness for use in benchmark review, and if needed, undertake additional data collection.
- If information from other sources is used, assess its comparability and undertake adjustments to make it useable for the review.
- If required information is not available, develop guidance on monitoring and reporting data that is needed for the review. Such guidance should include information on monitoring variables, frequency of reporting, verification protocol, and the acceptable data sources (such as metered readings, sale receipts etc.).

Data can be monitored using the scheme's monitoring framework or separately. The relevant authority can include benchmarking specific data needs in the overall MRV strategy of their policy instrument. Integrating data needed for review in the scheme's reporting mechanism can reduce efforts required at the time of the benchmark review. This includes situations where data on key benchmarking variables is either not reported or reported at a level of granularity not used for benchmark review (e.g., reporting occurs at the facility level while benchmarks are defined at sub-facility level).

Alternatively, the data collection approach can be separate from the instruments' overall MRV strategy. Data collection outside the MRV strategy can either use existing data sources, other planned data gathering or involve the gathering of new primary data.

- As an example of use of existing sources, Australia plans to use pre-existing data from their national reporting scheme for benchmark review. In such situations, it's important to check if the key variables in the data source are relevant and robust.
- In cases where the same set of covered entities come under the scope of more than one scheme, collecting information on the other scheme is also useful. For instance, in India's PAT scheme, information on the renewable purchase obligations (RPOs) of an entity is also expected to be reported.
- If none of these approaches are possible, a concerted monitoring and data collection effort for revising benchmarks would need to be planned. Consideration would need to be given to the length of time necessary to gather this new data and the costs of doing so.

Plan for data review

Data review will take the form of comparing the current benchmarks with the recalculated value using new and revised historical data. In order to recalculate the value, the newly collected data will need to be analyzed in the same manner as set out in *Step 3 Data Analysis*. Data quality and sufficiency will be analyzed and improved if necessary. Subsequently the new benchmark value can be determined.

7.3 Stakeholder engagement

Stakeholder engagement at this Step is required to both communicate and provide guidance on the ongoing monitoring and review process, and to consult them on the outcomes and potential updates.

Communicate and provide guidance on monitoring and review

Once the monitoring and review plan is established, it should be communicated to participating entities and investors, as this gives them time to align their long-term business strategies with the policy.

Because a key part of the monitoring and review plan is ongoing engagement with stakeholders for data collection, guidance on the process should be provided to stakeholders. Such guidance should include information on monitoring variables, frequency of reporting, verification protocol, and the acceptable data sources.

Consult stakeholders in the review and update process

Stakeholder involvement is important when updating benchmarks, both at the stage of reviewing data, and deciding whether to trigger and update.

During the data review, stakeholders can provide support in identifying where benchmarks become out of date or for the provision of new data. Jurisdictions generally adopt an open-door policy for receiving feedback from scheme participants which can inform the benchmark review. Most administrators accept feedback through emails, letters, or phone during the compliance period and review process. In addition, many jurisdictions undertake ex-post evaluations of the overall scheme and implementation experiences, and benchmark application impacts can also be received under this process

Once it is decided that benchmark amendment would be needed, a round of stakeholder consultation using the usual means of consultations (i.e., meetings, workshops and public input) is usually undertaken. For the EU ETS revision and impact assessment the Commission organized three stakeholder events throughout 2014 and two written consultations (a more specific one on free allocation and carbon leakage and a wider one regarding the options proposed for the ETS revision). Industry experts are also important stakeholders to engage with during the review and amendment of benchmarks, particularly whether the identified production variables are appropriate, on technological changes in the sector etc. An example of this comes from Australia, which has established a role for industry experts in the technical working groups that would be tasked with benchmark update for covered sectors.

Glossary

Activity parameter. Units of input (products, heat, services) or output (fuel, electricity consumed) which are used to measure the environmental performance of an entity.

Benchmark. A benchmark, as used in this guide, is an emission intensity metric applicable for a specific activity and group of peers.

Crediting thresholds. Crediting thresholds are defined as the levels of environmental performance a covered entity must achieve in order to earn credits. The term is often used to discuss baselines that are more conservative than business as usual baselines. Benchmarks can be used for determining crediting thresholds. In the context of this guide, a crediting threshold is synonymous to benchmarked crediting thresholds.

Entity. A stakeholder which is subject to the benchmarking exercise, and considered a single unit for the purpose of the exercise. In other contexts, used interchangeably with "facility." "installation," or "participant," where the last refers to stakeholder covered by a policy instrument that may involve benchmarking.

Environmental Performance. In the context of benchmarking for climate policy instruments, environmental performance is measured by a particular environmental impact (Greenhouse gas or CO_2 emissions, energy use, etc.) associated with a particular activity (production or outputs or consumption of inputs) performed by an entity in a peer group.

Emission reduction credits. Emission reduction credits are calculated as the difference between ex-ante estimated baseline emissions and actual emissions after implementation of an intervention in a defined timeframe.

Impact parameter. Units used to express the environmental impact of an activity, such as GHG emissions produced, or energy used.

Input based benchmark. Benchmarking tools typically involve comparison of emission intensity associated with particular practices. In input based benchmarking, this is done on the basis of the inputs to the production process—for example, the fuel, heat, technology or process—used to produce the product.

Output based benchmark. Benchmarking tools typically involve comparison of emission intensity associated with particular practices. In output based benchmarking, this is done on the basis of the output of the production process, typically the product produced.

Policymaker. Relevant public sector authority responsible for decision making on the benchmarking development process.

Reference values. Pre-existing emission intensity values which have been calculated for a specific sector and activity.

Relevant authority. A public-sector stakeholder with responsibility for executing the decisions of the policymaker, usually regarding the implementation of the benchmark development process.

Peers. Peers are entities belonging to the same group, from whom data is collected for the calculation of a benchmark. They must be sufficiently similar with regards to the parameters the benchmark is based on to be considered part of a peer group.

Representative. Used in the context of data representativeness, this is a measure of the quality of the data. Representativeness of data depends on whether the sample of responses received cover a proportionally diverse number of the peer group.

Sufficiency. Used in the context of data sufficiency in a data collection exercise, this is measured by the response rate as a proportion of the data requested in the collection exercise.

References

Australia, Department of Eenvironment. 2016. Draft Emissions Intensity Benchmark Guidelines for the Emissions Reduction Fund: Safeguard mechanism. Retrieved from http://www.environment.gov.au/system/files/consultations/493a2d2c-ce7d-443f-95d3-79b0c8a512d7/files/benchmark-guidelines.pdf

California Air Resources Board. July 2011. Cap-and-Trade Regulation section 95891(b), Appendix B: Development of Product Benchmarks for Allowance Allocation. Retrieved from http://www.arb.ca.gov/regact/2010/capandtrade10/candtappb.pdf

California Air Resources Board. March 2014. Cap-and-Trade Regulation section 95891(b), Appendix A: Additions and Amendments to Product-Based Benchmarks in the Cap-and-Trade Regulation. Retrieved from http://www.arb.ca.gov/regact/2013/capandtrade13/2appabenchmarks.pdf

California Air and Resources Board. September 2013. Cap-and-Trade Regulation section 95891(b), Appendix C: New and Modified Product-Based Benchmarks. Retrieved from http://www.arb.ca.gov/regact/2013/capandtrade13/capandtrade13isorappc.pdf

European Commission. 2016. Free allocation based on benchmarks - European Commission. Retrieved August 26, 2016, from https://ec.europa.eu/clima/policies/ets/allowances_en#tab-0-1

Ecofys and Fraunhofer Institute for Systems and Innovation Research. 2009. *Developing Benchmarking Criteria For Co2 Emissions, on behalf of DG ENV.*

Ecofys and The Green House. October 2014. *Emissions intensity benchmarks for the South African carbon tax- Technical support study.* Retrieved from http://www.treasury.gov.za/publications/other/GHG_Emissions_Intensity_Benchmarks_for_SA_Carbon_Tax.pdf

Ecofys, Fraunhofer Institute for Systems and Innovation Research, and Öko-Institut. 2009. *Methodology for the free allocation of emission allowances in the EU ETS post 2012 - Sector report for the cement industry.* Retrieved from https://ec.europa.eu/clima/sites/clima/files/ets/allowances/docs/bm_study-cement_en.pdf

Ecofys, Fraunhofer Institute for Systems and Innovation Research, and Öko-Institut. 2009. *Methodology for the free allocation of emission allowances in the EU ETS post 2012; Study consisting of a main report and 13 sector studies.* Retrieved from http://ec.europa.eu/clima/policies/ets/allowances_en#tab-0-2

Entec and NERA. 2005. *EU Emissions Trading, Scheme Benchmark Research for Phase 2 - Final Report for DTI.*

European Comission. (2011). *Guidance Document n°9 on the harmonized free allocation methodology for the EU-ETS post 2012; Sector-specific guidance.* Retrieved from https://ec.europa.eu/clima/policies/ets/allowances_en#tab-0-1

European Union. 2011. COMMISSION DECISION of 27 April 2011 determining transitional Union-wide rules for harmonised free allocation of emission allowances pursuant to Article 10a of Directive 2003/87/EC of the European Parliament and of the Council.

Greenhouse Gas Protocol. (n.d.). *GHG Accounting Standards.* Retrieved from http://ghgprotocol.org/standards

Greenhouse Gas Protocol. (n.d.). *Sector Tool Sets.* Retrieved from http://www.ghgprotocol.org/calculation-tools/sector-toolsets

The Green House and Ecofys. 2014. *Emissions intensity benchmarks for the South African carbon tax- Technical support study, for the World Bank.*

Hayashi, D., & Michaelowa, A. 2013. Standardization of baseline and additionality determination under the CDM. *Climate Policy*, *13*(2), 191–209. http://doi.org/10.1080/14693062.2013.745114

International Aluminium Institute. 2006. *Aluminium Sector Greenhouse Gas Protocol – addendum to WRI/WBCSD Greenhouse Gas Protocol.* Retrieved from http://www.world-aluminium.org/media/filer_public/2013/01/15/fl0000127.pdf

Ministry of Environment, New Zealand. 2016. Industrial allocation calculators | Ministry for the Environment. Retrieved August 25, 2016, from http://www.mfe.govt.nz/climate-change/reducing-greenhouse-gas-emissions/new-zealand-emissions-trading-scheme/industry-7

Ministry for the Environment, New Zealand. 2009. *Development of Industrial Allocation Regulations under the New Zealand Emissions Trading Scheme: Consultation Document.* Wellington.

Ministry of Power, India. 2015. *Normalisation document and Monitoring and Verification Guidelines-Cement Sector*. Retrieved from: https://www.beeindia.gov.in/content/pat-3

Perspectives Climate Change. (May 2010). Towards a more standardised approach to baselines and additionality under the CDM.

Partnership for Market Readiness. 2013. *Supporting GHG Mitigation Actions with Effective Data Management Systems.* Retrieved from https://openknowledge.worldbank.org/handle/10986/21828

Partnership for Market Readiness. 2015a. *Carbon Leakage: Theory, Evidence and Policy Design*. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/22785/K8516.pdf?sequence=1&isAllowed=y

Partnership for Market Readiness. 2015b. *Options to Use Existing International Offset Programs in a Domestic Context* (Technical Note). Partnership for Market Readiness.

Partnership for Market Readiness. 2016. *Greenhouse Gas Data Management Building Systems for Corporate/ Facility-Level Reporting.* Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/23741/K8658.pdf?sequence=5&isAllowed=y

Prag, Andrew and Gregory Briner. 2012. *Crossing the threshold: Ambitious baselines for the UNFCCC New Market-based Mechanism*. Retrieved from http://www.oecd.org/env/cc/50315387.pdf

Warnecke, C., H. Fekete, & T. Day. 2015. *Bilateral agreements as basis towards piloting sectoral carbon market mechanisms*. Retrieved from http://www.bmub.bund.de/fileadmin/Daten_BMU/Pools/Forschungsdatenbank/fkz_3712_41_507_carbon_market_mechanisms_bf.pdf

World Business Council for Sustainable Development. 2011. *Cement Sustainability Initiative - The Cement CO2 and Energy Protocol.* Retrieved from http://www.wbcsdcement.org/pdf/tf1_co2%20protocol%20v3.pdf

Wesselink, B., S. Klaus, A. Gilbert, & K. Blok. 2008. *The IFIEC method for the allocation of CO2 allowances in the EU Emissions Trading Scheme: a review applied to the electricity sector*. Retrieved from http://www.ecofys.com/files/files/pecsnl074036finalreportifiec-2_02.pdf

World Bank. 2016. *State and Trends of Carbon Pricing 2016*. Retrieved from http://www.ecofys.com/files/files/wb_report_2016_161018_screen.pdf

Annexes

A1. Benchmarking in the surveyed jurisdictions

This Guide was developed using an empirical approach, based on the evidence gathered through desk-based research of publically available information, and surveys of selected jurisdictions. These jurisdictions were chosen on the basis of instrument type and sectors, aiming for a broad coverage. The surveyed jurisdictions, and the instruments and sectors covered are presented in Table 14. The surveys were carried out in May and June 2016, using both questionnaires and follow up phone interviews.

Table 14: Summary of application of benchmarking and climate policy instruments in selected jurisdictions

Jurisdiction	Organization surveyed	Instrument and application	Implementation	Sectors	Further Information
Australia	Department of the Environment	*Safeguard mechanism;* Setting baselines for new entrants, similar to S-CP	Operational 2016	Mining, oil and gas, transport sectors.	Safeguard mechanism, which is part of the Emissions Reduction Fund
California (CA)	California Air Resources Board	*ETS;* Allowance distribution using benchmarks	Operational in 2013	Phase 1 (2013-2014): Electricity generation (including imports), Industrial sources Phase 2 (2015 onward): Sectors covered in Phase 1, plus distributors of transportation fuel, natural gas and other fuels	Allowance allocation under California Cap and Trade system
Cambodia		*CDM* Standardized baseline development	Adopted in 2013	Rice milling	UNFCCC secretariat's website
China (Shenzhen)	No response to questionnaire, interviewed	*ETS;* Carbon intensity benchmarks for allowance distribution	Implemented in 2013	Power, Water Supply, Manufacturing, Buildings, (Transport from public buses and taxis to potentially be included at a later date)	Capacity building for establishment of ETS in China
Colombia	No response to questionnaire, interviewed	*S-CP;* Baseline setting using benchmarks	Early design stage	Transport in MRP, but now focus more on buildings	Documents not available in public domain

Jurisdiction	Organization surveyed	Instrument and application	Implementation	Sectors	Further Information
European Union (EU)	European Commission - DG Climate Action	*ETS;* Allowance distribution using benchmarks	Implemented 2005	Industry, Power, Domestic aviation	EU ETS industrial allocation based on benchmark, and Aviation allocation based on benchmarks
India	GIZ-India	*EETS (Performance Achieve and Trade Scheme - PAT);* obligation distribution using benchmarks	Implemented in 2012	Industry and Power (PAT I), Railways (included along with other sectors in PAT II)	Website of Bureau of Energy Efficiency, the relevant authority for PAT Scheme
Japan	Institute of Global Environmental Strategies	*S-CP (Joint Crediting Mechanism – JCM);* Technology benchmarks in baseline setting	Implemented in 2013	Differ depending on partner country	JCM website
Tokyo	Tokyo Metropolitan Government	*ETS;* Allowance distribution using input based benchmarks	Implemented in 2010	Buildings, District heating / cooling	Tokyo cap and trade
Kazakhstan	Joint Stock Company Zhasyl damu	*ETS;* Allowance distribution using benchmarks	Implemented in 2013 (currently suspended)	Industry, Power	Kazakhstan – PMR Technical Partner
Mexico	No response to questionnaire, interviewed	*S-CP*	Early design stage	Urban, Transport, Refrigeration	Documents not available in public domain

Jurisdiction	Organization surveyed	Instrument and application	Implementation	Sectors	Further Information
New Zealand (NZ)	New Zealand Ministry for the Environment	*1. ETS;* Allowance distribution using benchmarks *2. Carbon Tax (Negotiated Greenhouse Agreements – NGA);* Setting tax free thresholds	ETS in 2010; NGAs in 2002	Industry, Power, Upstream (Buildings, Transport, Domestic aviation), Waste, Forestry	New Zealand ETS industrial allocations
Republic of Korea	*No response to questionnaire received*	*ETS;* Allowance distribution using benchmarks	Implemented in 2015	Industry, Power, Waste, Domestic Aviation, Buildings	Korean Emissions Trading Scheme
Sri Lanka	World Bank representative	*S-CP* (benchmarks not used) [29]	Design stage	Power sector	Documents not available in public domain
Tunisia	GIZ Tunisia	*S-CP* (To be defined)[30]	Design stage	Cement, Electricity	Documents not available in public domain
South Africa (SA)	National Treasury	*Carbon Tax;* Rebate distribution using benchmarks	Design stage	Industry, Transport, Commercial energy	Draft Carbon Tax bill, South Africa
UK	Ricardo Energy and Environment representative	*EETS (Climate Change Agreements-CCA);* Obligation distribution using benchmarks	1999	Industry, Agriculture	UK Climate Change Agreements

[29] Although a sectoral crediting instrument is being developed, Sri Lanka has decided not to use benchmarks for the time being. Nonetheless, a lot of their experience with this instrument is of relevance to this Guide, and therefore they are included here.

[30] At time of writing, Tunisia has yet to determine whether benchmarks will be developed. Nonetheless, their experience is of relevance and therefore included in the Guide.

A2. Example Data Collection Templates

Examples of the data collection templates used in the UK's Climate Change Agreements, and New Zealand's ETS collecting data on are provided below. UK's CCA template collects data on an entity's ("target unit") production levels and energy consumed. NZ's ETS template collects data on an entity's production levels, revenue and sales, and calculates emissions based on activity levels.

UK CCA: Relative Energy Reporting Template

A template in Excel spreadsheet format is available online. It includes an initial instructions sheet, explaining how to use the template. An example is provided below.

Instruction sheet: Absolute energy targets
This reporting template should be used by target units with absolute energy targets. Alternative templates are available for target units with other target types.
The template is divided into a number of sections. Instructions for completing each of these are provided below, and further guidance is given in comment boxes within the reporting template.
Section 1: Report Details
The table in this section should be completed to provide details on the report period and report version.
Section 2: Target Unit Details, Targets and Previous Performance
This section will be automatically populated using the data that we hold in the register for your target unit.
Section 3: Actual Target Period Performance for Target Facility
This section is used to report the actual performance of the target unit within the reporting period.
Data should be entered into the green cells for total production, as well as for all fuels used within the reporting period. The energy and throughput units used should match those shown in Section 2.

Entities must fill out the second sheet of the template, composed of five sections. In Section 2 of the template, the template automatically populates details of the installation concerned, targets and pervious performance, based on basic information provided in Section 1. In Section 3, they provide details of actual performance during the period.

Section 1: Report details

Sector	
Target Period	
Report Version	
Template Vers.	
Report Date	
Your Name	
Your Email	

Section 2: Target Unit details, target and previous performance

		Target Unit
TU Details	Identifier	
	TU Operator	
	No of facilities	
	Target Type	
	Energy Unit	
	Throughput units	
	Base year start date	
	Base year Energy ()	
	Base year Throughput ()	
TP Target	Value of latest agreement target ()	
	Value of latest agreement target %	

Previous Performance	Surplus CO_2 from previous Target Period (tonne CO_2)	

Section 3: Actual Target Period Performance for Target Facility

		Fuel Conversion Factors (tC/)	Target Unit	Target Unit Entry
Actual Target Period Performance (throughput and fuel split over 2 years)	Identifier		0	0
	Target Period (2 years) Total Production Units		0.000	
	Electricity used (PRIMARY) ()	0.0000546	0.000	
	Natural Gas used ()	0.0000505	0.000	
	Fuel Oil used ()	0.0000732	0.000	
	Coal used ()	0.0000794	0.000	
	Coke used ()	0.000117	0.000	
	LPG used ()	0.0000585	0.000	
	Ethane used ()	0.0000545	0.000	
	Kerosene used ()	0.0000673	0.000	
	Petrol used ()	0.0000643	0.000	
	Gas Oil/ Diesel Oil used ()	0.0000758	0.000	
	Naphtha used ()	0.0000646	0.000	
	Petroleum Coke used ()	0.0000908	0.000	
	Refinery Gas used ()	0.0000671	0.000	
	Other fuel - 01 - used ()		0.000	
	Other fuel - 02 - used ()		0.000	
	Other fuel - 03 - used ()		0.000	
	Other fuel - 04 - used ()		0.000	
	Other fuel - 05 - used ()		0.000	
	Other fuel - 06 - used ()		0.000	
	Other fuel - 07 - used ()		0.000	
	Other fuel - 08 - used ()		0.000	
	Other fuel - 09 - used ()		0.000	
	Other fuel - 10 - used ()		0.000	
	Other fuel - 11 - used ()		0.000	
	Target Period Total Energy		0.000	No data entered

Full details available at source: Climate Change Agreements: operations manual, UK Environment Agency, 2013, 2016 https://www.gov.uk/government/publications/climate-change-agreements-operations-manual--2

New Zealand ETS: Emissions Intense Trade Exposed – Industrial Allocation Data Collection Template

A template in Excel spreadsheet format is circulated by the relevant authority. It includes an initial instructions sheet, explaining how to use the template. An example is provided below

Emissions-Intense Trade-Exposed — Industrial Allocation Data Collection
1. This data template is to be used by entities to submit data in response to the Gazette Notice. The template has been designed in accordance with data requirements for assessing EITE activities as set out in the Climate Change Response Act 2002 and should be completed with reference to the Gazette Notice, Guidance materials, activity definitions and FAQs published on the Ministry for the Environment website.
2. This data template is to be used for submitting data that applies to each instance of an activity.
3. Entities are requested to provide data in all sheets using theGazette Notice and Guidance materials as the key reference. Data can only be entered into cells that are shaded light blue, however it is not necessary that all of these cells contain data.
4. Entities should only submit data into this template on the basis of final activity definitions as published in the Gazette Notice.

Entities must fill out basic information in the first sheet

1	Activity	Production of glass containers
2	Company Name	
3	Holding Account Number	
4	Facility Name (if applicable)	
5	Physical Address	
6	Contact Name	
7	Postal Address	
8	Phone	
9	Fax	
10	Email	

In the second sheet, historical data on Production, Sales and Revenue are requested.

ACTIVITY		Production of glass			
COMPANY NAME		0			
ADDRESS		0			
Output					
			Units produced of Basis of Allocation*		
Output name**		**Unit**	**2006/07**	**2007/08**	**2008/09**
11	Blown and pressed glass containers	tonnes			
			Activity Outputs Produced**		
Output name**		**Unit**	**2006/07**	**2007/08**	**2008/09**
12	Blown and pressed glass containers	tonnes			
Determination of Total Revenue by Actual Sales					
			Units Sold Externally***		
Output name		**Unit**	**2006/07**	**2007/08**	**2008/09**
13	Blown and pressed glass containers	tonnes			

In the third sheet, historical data fuel combustion or usage, electricity consumption, and consumption of other materials are used to calculate fuel combustion, electricity, steam, and industrial process emissions.

ACTIVITY	Production of
COMPANY NAME	0
FACILITY ADDRESS	0

FUEL COMBUSTION EMISSIONS				
Item	Sources	Emissions (t CO_2e)		
		2006/07	2007/08	2008/09
Solid fuels - Schedule 2, Table 2				
18	Coal - Lignite - Waimumu and Roxburgh fields	0	0	0
19	Coal - all other fields, or peat	0	0	0
20	Coal - Sub-bituminous	0	0	0
21	Coal - Bituminous	0	0	0
Item	Sources	Emissions (t CO_2e)		
		2006/07	2007/08	2008/09
Gaseous fuels - Schedule 2, Tables 4				
22	Natural gas - propane	0	0	0
23	Natural gas - butane	0	0	0
24	Natural gas - LPG (P60:B40)	0	0	0
25	Natural gas - LNG	0	0	0

Source: Provided by the New Zealand Ministry for the Environment - Manatū Mo Te Taiao.

WORLD BANK GROUP
Climate Change

PMR | *Pricing Carbon to Achieve Climate Mitigation*

http://www.thepmr.org
pmrsecretariat@worldbankgroup.org