

**DRAFT GUIDELINES FOR QUALITY ASSURANCE AND QUALITY CONTROL OF DATA USED IN THE ESTABLISHMENT OF STANDARDIZED BASELINES****(Version 01.0)****I. Background**

1. The Executive Board of the clean development mechanism (CDM) (hereinafter referred to as the Board), at its sixty-second meeting, approved the “Guidelines for the Establishment of Sector Specific Standardized Baselines” (hereinafter referred to as the SB Guidelines) to meet the request from the sixth session of the Conference of the Parties serving as the Meeting of the Parties to the Kyoto Protocol (CMP).
2. The Board at its sixty-third meeting adopted the “Procedure for submission and consideration of standardized baselines”. The Board requested the secretariat to prepare an implementation plan for the procedure for consideration by the Board at its sixty-fifth meeting at the latest. As the use of standardized baselines is voluntary, the Board stressed that the level of stringency of the standardized baselines should be sufficient to ensure a high level of environmental integrity. The Board also requested the secretariat to take into account this principle while developing the work programme to implement the SB Guidelines.
3. For the implementation of the SB Guidelines, the secretariat developed general guidelines to ensure the quality of data required for the establishment of the standardized baselines.

II. Introduction

4. The development of standardized baselines is a data-intensive process. The data required for quantifying baseline greenhouse gas (GHG) emissions in a particular sector include data on activities in a country or a group of countries such as energy usage, industrial production statistics, production technologies, demographic data, process-related characteristics and mitigation-related practices. Some data might already be available to Designated National Authorities (DNAs) (primary data) but in many cases it entails the process of data collection from data providers. The collection of data from multiple sources can lead to inconsistencies and therefore poses difficulties in ensuring data quality. The purpose of this document is to specify provisions and processes for ensuring data quality and to provide guidance on practical aspects of data collection to ensure quality of data, such as the use of sector-specific data templates.¹
5. This document aims to define best practices given the current state of scientific knowledge and data availability, which could help DNAs improve their institutional capacities for data management.
6. The best practice to ensure data quality as outlined in this document is twofold:
 - (a) Proactively² preventing potential risks that could cause quality deterioration, with a well-designed data management system, well-trained personnel and the culture of data quality; and

¹ Sector-specific data templates intended for use by DNAs to establish standardized baselines could reduce the risk of inconsistencies.

² It is the best practice to check the quality of data as early as possible. DNAs should avoid low-quality data being compiled in the datasets through pre- and post-submission quality checking.



- (b) Identifying and formulating data problems and implementing corrective actions, through regular reviews and continuous improvement processes.

III. Scope and applicability

7. The “Guidelines for the Quality Assurance and Quality Control of Data used in the Establishment of Standardized Baselines” (hereafter referred to as the QA/QC Guidelines) are applicable to those entities involved in the collection, processing, compilation and reporting of data needed for the establishment of sector specific standardized baselines. Such entities include:

- (a) DNAs that develop, validate, and/or own the datasets used for the establishment of standardized baselines;
- (b) Designated Operational Entities (DOEs)³ that conduct review procedures to assess the quality data used to establish the standardized baseline;
- (c) Project participants or other entities⁴ that develop standardized baselines.

8. The QA/QC Guidelines include the quality control (QC) procedures for compiling the required datasets and the quality assurance (QA) procedures for ensuring the overall quality of the datasets, based on data quality objectives and general provisions. The Guidelines also elaborate provisions for the documentation to be submitted in accordance with paragraph 8 of the “Procedure for Submission and Consideration of Standardized Baselines.”

9. The QA/QC procedures presented in this document are not exhaustive. DNAs could adapt the procedures to their own circumstances as long as best practices are applied.

IV. Key concepts and data quality objectives

10. For the purpose of this document, the following key concepts apply:

- (a) Data quality – data quality is a multi-dimensional concept and is commonly described as the degree to which data are “fit for use”.⁵ Data quality is ensured when it can be demonstrated that the datasets are relevant, complete, consistent, reliable, current, accurate and objective. In addition, processing data to derive standardized baselines should be conservative, secure, transparent and traceable;
- (b) Quality Control (QC) – QC is a system of routine technical activities to be conducted by a DNA to assess and maintain the quality of the datasets as the data are being compiled. It begins with pre-submission QC activities described in paragraph 23 below, followed by post-submission QC activities, internal review and a summary of the QC implementation (QC report);
- (c) Quality Assurance (QA) – QA is a system developed and implemented by a DNA to ensure that the QC system is designed to meet the data quality objectives below and the provisions specified in paragraphs 12 to 20 and

³ According to the “Procedure for Submission and Consideration of Standardized Baselines”, DOEs may prepare an assessment report. In the case where a Party (with fewer than 10 registered CDM project activities as of 31 December 2010) does not provide an assessment report in the first submission (allowed only once), the secretariat should prepare the assessment report.

⁴ The CMP at its sixth session decided that Parties, project participants, as well as international industry organizations or admitted observer organizations through the host country’s DNA may submit proposals for standardized baselines.

⁵ Knight, S. & Burn, J. (2005) Developing a framework for assessing information quality on the world wide web. *Information Science Journal* vol. 8. p.159-172.



implemented effectively. The conformity and the effectiveness of the QC system are reviewed by DOEs that are not directly involved in the compilation/development process of the datasets. Reviews are performed on the QC systems/processes used to establish the datasets and verify that objectives and all the provisions specified in the QA/QC Guidelines were met.

11. The following data quality objectives are intended to guide the implementation of the QA/QC procedures specified in paragraphs 21 to 36:

- (a) Relevance – select data and other information appropriate to the establishment of sector-specific standardized baselines. These are mainly activity data and information applicable for the determination of the baseline emissions in a sector;
- (b) Completeness – include all relevant activity data and information to produce the best representative standardized baselines in a sector. It requires procedures to avoid, identify and handle missing data (e.g. relevant companies that have been excluded or incomplete data entry);
- (c) Consistency – present all data in the same format and make the datasets compatible with other related data, which enables meaningful comparisons if necessary;
- (d) Credibility – identify and utilize credible data sources. Collected data/information should always reference their sources;
- (e) Currency – utilize the most recent data available in a sector in order to reflect the current economic and technological practices. The currency of the datasets is related to the data vintage and update frequency, which will be pre-determined for each sector;
- (f) Accuracy – reduce errors and uncertainties as far as is practical and cost-effective. The QC procedures should be well-designed to ensure the accuracy, which should focus on the procedures to avoid potential duplications and errors;
- (g) Objectivity – avoid biased, prejudiced and partial information. There should be no room for assumptions or differing interpretations in compiling the required data;
- (h) Conservativeness – ensure that any deviation from the QA/QC Guidelines that may lead to an overestimation of the baseline emissions in a sector is addressed by taking a conservative approach. Wherever other objectives in this document could not be met (e.g. there is a lack of accuracy), a conservative approach should be applied. For example, [the average values of the top 20%, based on performance data in the sector⁶ or] pre-determined conservative default values should substitute for missing, incomplete, invalid, old or incorrect data;
- (i) Security – develop procedures for restricted access to the datasets and maintain the security of the datasets. The procedures should include how to identify, process, present and manage confidential data. The data should be open to the public in an anonymous aggregate form;
- (j) Transparency – disclose sufficient and appropriate data and processes to allow monitoring of the quality of the compiled datasets and the generated outcomes.

⁶ Where applicable, the performance data may be derived from countries in similar conditions (economic, technical and environmental).



Through public engagement, reviewers may make decisions with reasonable confidence;

- (k) Traceability – document all data sources as well as measurement, calculation and estimation methods, which will enable the reproduction or review of the data used for the development of the standardized baselines by a third party. All documentation provisions for the review will be addressed in section VIII.

V. General provisions

Sector-specific data templates

12. To ensure the consistency of the datasets, DNAs should use sector-specific data templates provided by the secretariat. The data templates include standardized data formats and lists of relevant data, based on sector-specific situations and mitigation measures.

13. DNAs may request a revision to these data templates or propose a new data template [in accordance with the Procedures].⁷ If requested, further clarifications on the use of the data templates will be provided by the secretariat.

Data vintage and update frequency

14. DNAs should complete the data templates, based on the most recent data available using the vintage of data that is sector specific and defined by the Board. For a Party with fewer than 10 registered CDM project activities as of 31 December 2010, the Board may establish different data vintage for certain sectors if necessary.

15. DNAs should update the datasets used for the establishment of standardized baselines at a frequency that is sector-specific and defined by the Board. When updating the datasets, DNAs should apply a consistent approach. For example, data delivery protocol, compiling methods, assumptions and relevant calculations should be documented and comparable over time.

Completeness

16. DNAs should include all installations/facilities/companies (population) that contribute to the production of the sector output in a country or a group of countries, based on the scope of a sector defined in the data templates. Where some data (e.g. for companies) are dropped out due to non-response, DNAs should determine the level of population incompleteness by calculating the ratio of the number of dropped-out companies to the total number of companies and also by estimating the rate of the historical contribution of these dropped-out companies to the total production of the sector output. DNAs should apply a conservative approach to address this incompleteness (e.g. by assuming that top-performing companies are dropped) and provide justification for the conservative approach.

17. DNAs should provide all required data specified in the data templates. Incomplete data entries will be filled automatically by conservative default values if the values are pre-determined in the data templates. [Otherwise, DNAs should apply the average values of the top 20% performance data as conservative default values.] If other default values are applied, DNAs should demonstrate that those values are conservative.

⁷ The secretariat will develop further procedures for submission and consideration of sector-specific data templates for the establishment of standardized baselines.

*Accuracy*

18. DNAs should implement measures to avoid errors and duplications. Sources of uncertainties should be identified and reduced as far as practicable. However, where a high level of accuracy cannot be achieved, a conservative approach should be taken and a justification should also be provided.

19. When a sampling approach is employed to derive activity data, DNAs should apply statistically sound sampling approaches and comply with the sampling requirement for reliability [95/5 confidence/precision for industrial sectors and 90/10 confidence/precision for the others] and select sample sizes in accordance with the “Standard for Sampling and Surveys for CDM Project Activities and Programme of Activities”.

Transparency

20. Where applicable, DNAs should invite stakeholders to provide comments on the sector-specific standardized baseline and the process used to establish it. The consultation should be done in an open and transparent manner, in a way that facilitates comments from stakeholders and allows a reasonable time for comments to be submitted. DNAs should prepare a public consultation report that includes a summary of the comments provided by stakeholders and how they considered all the comments received.

VI. Quality control

21. DNAs should develop a QA/QC system that outlines QA/QC activities, processes, schedule and responsibilities of the personnel involved as well as the institutional arrangement. The goal of this system is to ensure that the above provisions and data quality objectives are met.

22. As part of the QA/QC system, DNAs should develop QC procedures that address how to ensure the data quality and apply a conservative approach to the whole QA/QC system. The QC procedures may be revised where necessary.

23. It is very important to facilitate the collection of high-quality data, so DNAs should conduct a pre-submission quality check by establishing a “data delivery protocol⁸” for data providers that describes specific rules and procedures for the collection and delivery of the requested data. The data delivery protocol could consist of the following components:

- (a) Purpose of data collection – DNAs could set multiple purposes in addition to the establishment of standardized baselines for a certain sector;
- (b) Data types – DNAs should specify the lists, units and formats of data requested, based on the sector-specific data templates, which will help to achieve a high level of consistency and relevance. Clear definitions or descriptions of the data provisions related to accuracy and/or conservativeness should be provided in an objective manner;
- (c) Data acquisition procedures – DNAs should provide clear guidance on data compilation and aggregation, for example how specific unit-level (equipment/facility) data should be consistently aggregated into entity-level (company) data. The procedures should indicate that data should be collected from valid data sources and only the current data should be compiled;

⁸ The protocol will help data providers better understand the requested data and ensure consistency, credibility, currency and in particular accuracy and objectivity of the data delivered by the data providers.



- (d) Credibility of data – data sources, references and responsible persons should all be documented by data providers;
- (e) Delivery requirements – DNAs should specify a scheduled time frame and types of the deliveries. DNAs could request periodical reports if necessary. It is recommended that data providers prepare a summary report explaining how the data was collected and how the quality of the data was ensured and including a declaration of conformity. This summary report should include all identified issues related to data quality (e.g. intrinsic uncertainty, limited data availability or inconsistent data system). If a conservative approach has been used to address the data quality objectives that could not be achieved, this conservative approach should be described in the summary report;
- (f) Confidential issues – DNAs should describe how they will address issues related to confidentiality;
- (g) Contact – DNAs should be responsive to queries from data providers and provide assistance whenever requested, so the protocol should contain contact information and define the deadline for responses to queries. DNAs should establish lines of effective communication and feedback with data providers to identify specific opportunities to improve the quality of data.

24. The data delivery protocol should be distributed to the entire target population of data providers. DNAs should define the scope of the population in a particular sector, based on the definition of the “sector” established by the secretariat. DNAs should ensure that there is no duplication or missing entities from the population. For example, a company included in a sector “A” should not be included in another sector “B” unless the data provided by the company for each sector are not interrelated and double counting does not occur.

25. DNAs should conduct a post-submission quality check before compiling into the data template. That is, DNAs should approve the submitted data after assessing the credibility of the data sources and the accuracy of the data, based on DNAs’ primary data and reliable secondary data (e.g. international data or academic data) as well as documents submitted by data providers. Where necessary, DNAs should ensure the credibility of the data by arranging for a third party to conduct a check of the facilities/companies. The following should be checked:

- (a) Whether all relevant facilities and all the relevant activities within the facilities have been taken into account;
- (b) Whether the data are current and meet the data vintage provision;
- (c) Whether reliable data systems are established and how the systems are operated and maintained (logs of operation and maintenance of the systems may be required);
- (d) Whether all data sources are documented accurately;
- (e) Whether the referenced materials are available;
- (f) How the data source quality has been assessed (e.g. calibration frequency or the accuracy of measuring instruments);
- (g) Whether the level of aggregation is appropriate and consistent with the data acquisition procedures specified in the data delivery protocol;
- (h) Whether the data processing (calculation/estimation) is traceable and appropriate if the data are derived from a different procedure from the procedures described in the data delivery protocol;

**CDM - Executive Board**

- (i) Whether the data are consistent with official and publicly available statistics or the DNAs' primary data (cross-checking wherever applicable). Discrepancies do not necessarily imply that the datasets are erroneous. However, such discrepancies should be justified or explained;
 - (j) Whether the sampling requirement is met if the data are derived from sampling, including a review of the representativeness of the sample and the appropriateness of the methods used for scale-up.
26. When certain problematic data are identified as per paragraph 25 above, DNAs should implement the following corrective actions before approval:
- (a) Request data corrections. In order to obtain correct data, DNAs could provide assistance to improve data systems and management practices for data providers;
 - (b) Apply a conservative approach as specified in the QA/QC system;
 - (c) Replace with DNAs' primary data and reliable secondary data, using conservative approaches. In cases where DNAs utilize activity data collected for other purposes with different systems and approaches, DNAs should confirm that the official national statistics offices or other data compiling agency have performed reliable QC activities.
27. The approved data are the input data for the data template. When compiling the data in the data template, DNAs should identify whether the following risks exist and take appropriate actions to prevent or solve them through internal review:
- (a) Double counting or duplication of records – the data approved per each data provider should be included only once in the data template (one spreadsheet per company);
 - (b) No response – DNAs should encourage the active participation of data providers and be responsive to their needs. For any non-responses, a conservative approach should be applied to replace unavailable data;
 - (c) Limited data availability – due to poor quality data or data compiling systems, some data may not be approved by DNAs. DNAs should apply a conservative approach to these unqualified data. DNAs should provide data providers with assistance to improve data systems and management practices for future data collection;
 - (d) Incorrect data entry – great care should be taken to avoid typographical errors, erroneous entry (in a wrong column or a wrong category) and duplication of entries (within one spreadsheet);
 - (e) Incomplete data entry – in order to meet the completeness provision, all mandatory fields in the data templates should be filled in. If there is an incomplete entry, a conservative approach should be applied;
 - (f) Low-quality data processing – If further consolidation, calculation or conversion is required, DNAs should establish an additional step to double check the outcomes (e.g. recalculation or reconciliation). Manual processing of data should be avoided and the use of spreadsheet is preferred. The formulae used should be described;
 - (g) Inconsistency – DNAs should ensure that the datasets in the data template are consistently compiled. Consistent categories, methods, processes and approaches should be applied. When updating, consistency should be checked as specified in paragraph 15.



28. The data templates automatically generate aggregated outcomes for standardized baselines, but DNAs could cross check the outcomes against a number of similar international datasets for economic indicators or against the relevant primary data or historical data if available.

29. The level of uncertainties depends on the data availability and knowledge of underlying process and inference methods. DNAs should identify key causes of uncertainties, such as a lack of completeness, limited data availability, missing data, misclassifications, non-systematic process of collecting data and misreporting. DNAs should quantify such uncertainties⁹ and take corrective actions to address them (e.g. by applying a discounting factor).¹⁰

30. DNAs should document (in a QC report) how the QC procedures were implemented and how the data quality objectives and the general provisions were met. The QC report should contain how all evidence/references to data sources were checked. It should include justifications on the selected approach for obtaining reliable input data (e.g. measurement, calculation, national statistics, sampling, surveys and/or measurement campaign). Information on the uncertainties associated with activity data and major issues regarding the quality of input data, methods, processing or estimates should be addressed in this report. If updating, the report should highlight changes in data inputs or methods, substantial divergences in the datasets and a trend analysis if necessary. The QC report should summarize key findings and present a plan for how to address any identified major issues in the future (e.g. training for data maintenance personnel or automatic data management systems).

VII. Quality assurance

31. The QA procedures should be based on an approach for assessing the quality of the data management system rather than checking the accuracy of a specific set (or sets) of data.

32. As part of the QA/QC system, DNAs should develop QA procedures for the systematic identification, formulation and analysis of risks of not meeting quality objectives/provisions for the datasets and for defining and implementing activities that mitigate the identified risks. The QA procedures should include update plans to continuously improve the accuracy of the datasets and the efficiency of the overall data system. DNAs could develop procedures to secure official organizational approval and support for the datasets, if necessary.

33. As part of the QA system, the DOE contracted by the DNA should assess the system put in place to ensure that the QC was implemented and performed effectively. This requires assessing the QC system against the data quality objectives and general provisions established in this document. It also includes assessing whether the QC system has been implemented as designed.

34. During the review of the QC system by the DOE, it is recommended to include experts in relevant technical fields in the assessment team to check whether the QC system, procedures and approaches as well as the final standardized baselines are reasonable.

35. The elements of the QC system that are to be assessed by the DOE include:

- (a) System availability – identify whether a “standardized” data system (collection, consolidation and maintenance) is currently in place and a procedure for

⁹ Refer to the 2006 IPCC report for further guidance on uncertainty <http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/1_Volume1/V1_3_Ch3_Uncertainties.pdf>.

¹⁰ Such as discounting factors as stated in FCCC/SBSTA/2003/10/Add.2, p 25.



reporting activities conducted as part of the QC system has been developed and implemented;

- (b) Conformity – assess whether the QA/QC system, the QC procedures and all the approaches to develop the datasets met the data quality objectives and the general provisions specified in paragraphs 12 to 20. In particular, DOEs should assess whether a conservative approach has been applied in a consistent manner and whether the credibility of the data sources has been ensured. DOEs should check whether the QC procedures (i) were developed in accordance with the provisions set in the QA/QC Guidelines, and (ii) were effectively implemented within the QA/QC system. The data delivery protocol should be assessed to determine whether it was consistent with the sector-specific data template. DOEs may also review the summary reports by data providers and assess whether they were generated in accordance with the data delivery protocol. DOEs should assess whether the transparency provision was met, based on the public consultation report and the QC report;
- (c) Traceability – check whether all data and information (e.g. data sources) relating to the datasets and procedures for standardized baselines were clearly documented;
- (d) Responsiveness – evaluate whether the data delivery protocol met the provisions of the QA/QC Guidelines and whether communication with data providers was timely and efficient;
- (e) Adaptability – assess whether DNAs addressed the major issues identified and whether the system through its procedures were modified accordingly;
- (f) Security – check whether a security system for data management is in place and has operated effectively. Identify whether any issues related to security occurred;
- (g) Error tolerance – check whether DNAs planned to minimize errors and established and implemented procedures to identify and correct errors proactively.

36. The results of the QA activities should be documented and included in an assessment report by the DOE. The DOE should prepare a summary of findings including the key issues identified and provide the overall evaluation. If necessary, DOEs could provide recommendations to improve the quality of data or the overall data system. The assessment report should be made available to DNAs. DNAs may provide responses to the findings in the assessment report and/or implement appropriate corrective actions that address the findings. The responses as well as the corrective actions implemented should be well documented in the final assessment report.

VIII. Documentation provisions

37. DNAs should document and maintain all data and information relating to the establishment of standardized baselines and submit the datasets compiled in the sector-specific data templates and the following documents in accordance with the “Procedure for Submission and Consideration of Standardized Baselines”:

- (a) QA/QC system – as specified in paragraph 21;
- (b) Data delivery protocol – as specified in paragraph 23;
- (c) Summary reports – as specified in paragraph 23 (e), data providers should prepare these reports. However, if the reports are not submitted, DNAs should produce these reports during approval;



CDM - Executive Board

- (d) Raw data and any supplementary documents submitted by data providers;
- (e) Primary data – wherever used as reference or raw data for the establishment of standardized baselines, the data sources and all the issues related to data quality of the primary data (e.g. the approach of data collection, processing and treatment of the associated uncertainties) should be documented;
- (f) Secondary data – wherever used as reference or raw data for the establishment of standardized baselines, the data sources and all the issues related to data quality of the secondary data should be documented;
- (g) Datasets compiled in the data templates;
- (h) Public consultation report – as specified in paragraph 20;
- (i) QC report – as specified in paragraph 30;
- (j) Assessment report – as specified in paragraph 36 above.

38. Secure archiving of complete datasets is important. DNAs should retain all the data/information for a period of five years after the submission of the standardized baselines. Great care should be taken to ensure the confidentiality. DNAs should have a secured data maintenance system including setting codes set for access control, strategies for unexpected damage or loss of data and procedures to protect confidential data.

- - - - -

History of the document

Version	Date	Nature of revision(s)
01.0.	8 December 2011	Initial publication. Prepared by the secretariat as requested by the Board (EB 65, paragraph 96, 25 November 2011).
Decision Class: Regulatory Document Type: Guideline Business Function: Methodology		