Ref: ISO 19157-1:2023

# Guidelines

## for Bhutan Standard of Data Quality

2024

Department of Survey and Mapping
National Land Commission Secretariat
Royal Government of Bhutan

# GUIDELINES FOR GEOGRAPHIC INFORMATION DATA QUALITY

(Ref: ISO 19157-1:2023)

November 2024

National Land Commission Secretariat
Royal Government of Bhutan

Centre for Geo-Information, National Land Commission Secretariat
203 Drophen Zur Lam 15 NW, Kawang Lam, Jangsa
Thimphu - 11001



## Technical Working Group

1. Samdrup Dorji, Chief Survey Engineer, Geo-Informatics Division
2. Lobzang Tobgye, Superintendent Survey Engineer, Topographical Survey Division
3. Prakash Rai, Senior Survey Engineer, Geo-Informatics Division
4. Pema Wangda, Senior GIS Officer, Geo-Informatics Division
5. Sangay Nidup, GIS Officer, Geo-Informatics Division
6. Jigme Namgay, Survey Engineer, Geo-Informatics Division
7. Chencho Tshering, Survey Engineer, Geo-Informatics Division
8. Members, NSDI Stakeholder Agency, RGOB

# Contents

# Introduction

Data Quality refers to the degree of excellence in a product and services that satisfy the given objective. The quality of spatial data can be defined as its fitness for use. In other words, completeness of attributes in order to achieve the particular task can be termed as data quality. Data that is appropriate for use with one application may not be fit for use with another. Different users have different perceptions as to the importance of data quality.

Geographic data is increasingly being shared, interchanged and used for purposes other than the producer's intended objective. Various agencies including private entities create the geospatial data and they need to assess the data quality standards in order to obtain better results. Geospatial data are created from various sources with different techniques which can have discrepancies in terms of resolution, precision and accuracy. Therefore, the main purpose of describing the quality of geographic data is to facilitate the users to assess and make a choice of the dataset best suited to application needs or requirements.

Geospatial data may not be perfectly accurate, but it can be useful to some extent given applications and purposes; In other words, it has a quality (fit-for-use). To determine data quality, errors should be properly documented (i.e. data quality). It is essential that the results of the data quality are expressed in a comparable way and follow a common understanding of the data quality measures in ensuring the integrity, reliability and usefulness of data in various applications. These data quality measures assess different aspects of the data to ensure it is suitable for its intended use.

The issue of data quality is continuing to challenge the spatial data community. When used in GIS analysis, a dataset's quality significantly affects confidence in the results. Unknown data quality leads to uncertain decisions, increased liability, and loss of productivity. Conversely, decisions based on data of known quality are made with greater confidence and are more easily explained and defended. The concern for spatial data quality has increased in recent years due to a number of factors; Increased data production by the private sector and non-government

agencies, which are not governed by uniform quality standards (production of data by national agencies has long been required to conform to national accuracy standards). Increased use of GIS for decision support, highlighting the implications of using low quality data. Increased reliance on secondary data sources with proper quality assurance.

ISO 19157 of geographic information on data quality provides a framework for defining the quality of geographic data. The framework includes principles of evaluating quality, a conceptual model for handling quality information, a structure and content of data quality measures, and guidelines for reporting a quality evaluation.

## Scope

This guideline describes the general requirements for describing the quality of geographic data. It includes the components for describing the data quality, specifies the components and content structure of data quality measures, describes the general procedures for evaluating the quality of geographic data and establishes principles for reporting data quality. This document does not attempt to define minimum acceptable levels of quality for geographic data.

## Objective

The main objective of this guideline is to provide a basic required information for defining the quality of geographic data based on the Bhutan Standards (BTS) for data quality to the data producers.

## Normative Reference

- ISO 19157-1:2023 Geographic information — Data quality — Part 1: General requirements
- Guidelines for Data Product Specifications (NLCS, 2024)
- Guidelines for Metadata (NLCS, 2024)

# Terms and Definitions

Following terms and definitions are used in this document.

- **Accuracy:** Closeness of agreement between a test result or measurement result and the true value.
- **Conformance:** Fulfillment of a requirement.
- **Conformance Quality Level:** Threshold values for data quality results.
- **Correctness:** Correspondence with the data quality requirements.
- **Data Product Specification:** Specification of data product together with additional information that will enable it to be created, supplied to and used by another party.
- **Data Quality:** Degree to which a set of inherent characteristics of data fulfills its requirements.
- **Data Quality Unit:** Combination of a scope and data quality elements.
- **Metaquality:** Information describing the quality of data quality.
- **Data Quality Elements:** Components describing a certain aspect of the data quality.
- **Data Quality Measure:** Evaluation of data quality elements.
- **Data Quality Evaluation Procedure:** Methods used in applying and reporting quality evaluation methods and their results .
- **Data Quality Results:** Value or set of values resulting from applying data quality measures.
- **Universe of Discourse:** View of the real world that includes everything.
- **Data Quality Scope:** Specifies the extent, spatial and/or temporal, and/or common characteristic(s) that identify the data on which data quality is to be established and evaluated.

# Acronyms and Abbreviations

- **ISO:** International Organization for Standardization
- **DPS:** Data Product Specification
- **DQ:** Data Quality

# General Requirements for GI Quality

Data quality concepts provide an important framework for data producers, as well as for data users. It establishes a framework to identify the relevant aspects of spatial data quality through quality elements, establish well-described quality assessment measures, methods on how to express quality and appropriately report the results.

# Components of Data Quality

Geospatial data has different components to its quality of data. International Organization for Standardization (ISO) 19157 describes overall content of the data quality including data quality unit, scope of the data quality unit and data quality elements. These components are used to describe the overall data quality elements.

**Data Quality Unit**

A data quality unit is the combination of a scope and data quality elements.

**Scope of Data Quality Unit**

The scope of the data quality unit(s) specifies the set of data (feature class) based on the common characteristics like extent, spatial and/or temporal, on which data quality is to be established and evaluated.

**Data Quality Elements**

Data quality elements is a component describing a certain aspect of the quality of geographic data required for a dataset to be used and accurate. Data quality can be characterized by the data quality elements defined in Table 1.

- Completeness

Completeness refers to the presence and/or absence of features, their attributes and relationships. Completeness should mainly be used on the feature instance, describing whether the features in the universe of discourse are found in the dataset or not. However completeness can also be relevant for feature properties such as “attribute completeness” and “relationship completeness”. Before using

completeness for this, the logical consistency/conceptual consistency should be carefully considered.

- Logical Consistency

Logical consistency is defined as the degree of adherence to logical rules of data structure, attributes and relationships (data structure can be conceptual, logical or physical) as defined by an organization in their Data Product Specifications.

- Positional Accuracy

Positional accuracy refers to the accuracy of the position of spatial data in a reference system.

- Temporal Quality

Temporal accuracy refers to the correctness of the temporal references of an item (reporting of error in time measurement).

- Thematic Quality

Thematic accuracy refers to the thematic aspects of spatial data considering the classification correctness, non-quantitative (qualitative) attribute correctness, and quantitative attribute accuracy.

Table 1: Data Quality Elements and Its Definition

| Data Quality Groups | Data Quality Element | Definition |
|---|---|---|
| Completeness | commission | excess data present in a dataset |
| | omission | data absent from a dataset |
| Logical consistency | conceptual consistency | adherence to rules of the conceptual schema |
| | domain consistency | adherence of values to the value domains |
| | format consistency | degree to which data is stored in accordance with the physical structure of the dataset |
| | topological consistency | correctness of the explicitly encoded topological characteristics of a dataset |

| Data Quality Groups | Data Quality Element | Definition |
|---|---|---|
| Positional accuracy | absolute or external accuracy | closeness of reported coordinate values to values accepted as true in a standard coordinate reference system |
| | relative or internal accuracy | closeness of the relative positions of features in a related dataset to their respective relative positions as true in a local coordinate reference system |
| | gridded data position accuracy | closeness of gridded data spatial position values to values accepted as true |

## Descriptor of Data Quality Element

Data quality elements and their descriptors are used to describe how well a dataset meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information. It is described by a reference to a quality measure, an evaluation method, quality result and a metaquality.

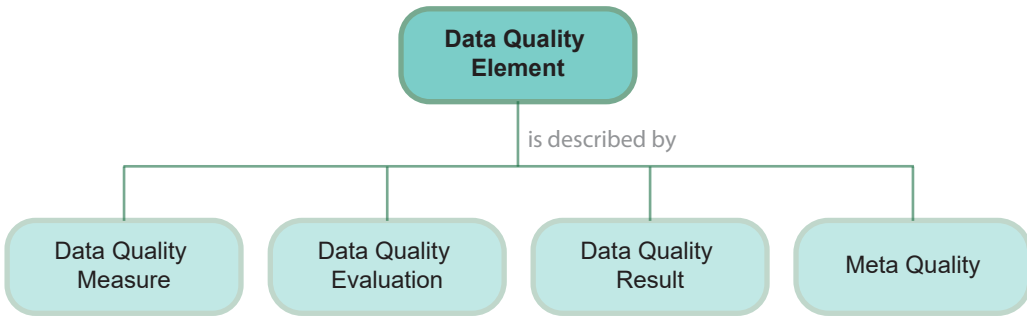



Figure 1: Data Quality Element Descriptor

**Data Quality Measure**

Data quality measures provide descriptors of the quality of geographic data through comparison with the real world features as described by the Data Product Specification or user requirements. Table 2 highlights the obligation/conditional component of data quality measures that needs to be taken into consideration for any data quality element.

Table 2: Components For Defining Data Quality Measures

| Component | Description | Obligation/condition |
|---|---|---|
| Name | Name of the data quality measure applied to the data | M |
| Data Quality Element Group | Name of the data quality element group for which quality is reported | M |
| Data Quality Element | Name of the data quality element for which quality is reported | M |
| Definition | Definition of the fundamental concept for the data quality measure | M |
| Description | Description of the data quality measure, including all formulae and/or illustrations needed to establish the result of applying the measure | C/if the definition is not sufficient for the understanding of the data quality measure concept |
| Data Quality Value Type | Value type for reporting a data quality result | M |

The list of data quality measures for each data quality element are described in Annexure I.

**Data Quality Evaluation**

The data quality evaluation is a sequence of steps to produce a data quality result. The process involved in the data quality evaluation is as described in Table 3.

Table 3: Process Steps Involved In Quality Evaluation

| Process Step | Action | Description |
|---|---|---|
| 1 | Specify data quality unit(s) | A data quality unit is composed of a scope and quality element(s). All data quality elements relevant to the data for which quality is to be described should be used. |
| 2 | Specify data quality measures | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided. |
| 3 | Specify data quality evaluation procedures | A data quality evaluation procedure consists of applying one or more evaluation methods. Including:<br>-standardized assessment method including conformance levels;<br>-sampling strategy |
| 4 | Determine the output of the data quality evaluation | A result is the output of applying the evaluation. |

- Data Quality Evaluation Methods

A data quality evaluation procedure comprises one or more data quality evaluation methods. Data quality evaluation methods can be divided into two main classes:

i. Direct

Direct evaluation methods determine data quality through the comparison of the

data with internal and/or external reference information. The direct evaluation methods are further sub classified by the source of the information needed to perform the evaluation, i.e. internal or external.

Internal direct data quality evaluation uses only data that resides in the dataset being evaluated. External direct quality evaluation requires reference data external to the dataset being tested.

For both external and internal direct evaluation methods, one of the following inspection methods may be used:

» full inspection: data quality evaluation tests every item specified by the data quality scope;
» sample-based inspection: data quality evaluation is performed on subsets of the geographic data defined by the data quality scope.

The full inspection is preferable. However, if the evaluation is based on a sample-based inspection the sample strategy should follow based on the ISO 19157-1:2023 Annex E Sampling methods for evaluating data quality.

ii. Indirect

An indirect evaluation method is a method of evaluating the quality of a dataset based on external knowledge or experience of the data product and can be subjective. This external knowledge may include information about the dataset's usage, lineage and purpose.

**Data Quality Results/Reporting**

A data quality result is provided for each data quality element. This can be a quantitative result, a conformance result, or a descriptive result.

- Quantitative Result

A quantitative result may be a single value or multiple values, depending on the values of attributes valueType and valueStructure defined in the description of the measure applied. While capturing the data results in quantitative format the following components of qualitative result must be recorded;

i) value: Record
ii) valueUnit: Unit of Measure
iii) valueRecordType: Record type

Example;
value (number of excess item) = 5, valueUnit = count, valueRecordType=integer

- Conformance Result

Conformance result is the outcome of comparing the value or set of values obtained from applying a measure to the data quality scope with the data product specification. For conformance result, the following components of conformance result must be recorded;

i) pass: Boolean( Yes/No)
ii) Specification: CI_citation
iii) Explanation: CharacterString

Example;
Pass: Yes, Specification: DPS for Fundamental Geospatial data (Scale 1/25,000), Explanation: All the data specified in the scope conforms to the format consistency specified in DPS.

- Descriptive Result

In some cases (e.g. with thematic and geo-scientific observations), it is not possible to produce a quantitative result for a data quality element. A subjective evaluation of an element can then be expressed with a textual statement as a data quality descriptive result.
Example;
The relative positional accuracy is higher between a geological feature and a nearby feature from a base map (roads, rivers, lakes etc.) than the absolute positional accuracy on the geological feature itself.

**Metaquality**

Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. These metaquality can also be reported through the metadata information as optional.

To ensure quality is achieved to the highest standard, geospatial data should be assessed and evaluated based on the defined data quality elements. The process of assessing the data quality is covered in section 8.2 of this document. Every geospatial producer needs to report the data quality based on the data quality reporting format reflected in Table 4.

Table 4: Data Quality Reporting Format

| Data Quality Components | Description |
|---|---|
| Quality Requirement | Define quality requirement of the particular dataset |
| Data Quality Element Group | Specify the particular data quality element-group that need to be assessed |
| Data Quality Element | Select or specify the data quality element |
| Data Quality Scope | Specifies the dataset extent that identifies the data on which data quality is to be established and evaluated. |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided.<br><br>**Name:** Name of the data quality measure applied to the data<br><br>**Definition:** Definition of the fundamental concept for the data quality measure<br><br>**Description:** Description of the data quality measure, including all formulae and/or illustrations needed and procedures to establish the result of applying the measure |

| Data Quality Components | Description |
|---|---|
| | **Data Quality Value Type:** Define the value type for reporting a data quality result (integer, real, boolean variable (true or false), string, date, percentage, ratio) |
| Data Quality Evaluation Procedure | Specify the data quality evaluation methods such as direct or indirect methods |
| Conformance Quality Level | Specify and set the threshold values that needs to conform the data quality results |
| Data Quality Result | The data quality results can be a quantitative result, a conformance result, a descriptive result or a coverage result. |

## Reference


- ISO 19157:2013 Geographic information — Data quality, Annex D List of standardized data quality measures

# Annexure I: List of Data Quality Measures

## 1. Completeness

In order to achieve well defined and comparable quality information, it is strongly recommended to carry out the evaluation and reporting of data quality using these data quality measures. Due to the nature of quality and geospatial information, this list cannot be complete. Therefore, users can come up with their own measure as described in this document if measures here do not meet the requirement.

i. Commission

Table 1: Number of Excess Items

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of excess items |
| 2 | Alias | – |
| 3 | Data quality element | Completeness |
| 4 | Data quality sub element | Commission |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Number of items within the dataset that should not have been in the dataset |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |

| | | |
|---|---|---|
| 12 | Example | – |
| 13 | Identifier | 1 |

Table 2: Number of Duplicate Feature Instances

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of duplicate feature instances |
| 2 | Alias | – |
| 3 | Data quality element group | Completeness |
| 4 | Data quality element | Commission |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Total number of exact duplications of feature instances within the data |
| 7 | Description | Count of all items in the data that are incorrectly extracted with duplicate geometries |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | Features with identical attribution and identical coordinates: two (or more) points collected on top of each other; two (or more) curves collected on top of each other; two (or more) surfaces collected on top of each other. |
| 13 | Identifier | 2 |

ii. Omission

Table 3: Number of Missing Items

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of missing items |
| 2 | Alias | – |
| 3 | Data quality element group | Completeness |
| 4 | Data quality element | Omission |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all items that should have been in the dataset and are missing |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 3 |

**2. Logical Consistency**

i. Conceptual Consistency

The data quality measures for the data quality element conceptual consistency are provided below.

Table 4: Number of Items Not Compliant With The Definitions In The DPS

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of items not compliant with the definitions in the DPS |
| 2 | Alias | – |
| 3 | Data quality element | Logical consistency |
| 4 | Data quality sub element | Conceptual consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all items in the dataset that are not compliant with the definitions in the DPS |
| 7 | Description | If the conceptual schema explicitly or implicitly describes rules, these rules shall be followed. Violations against such rules can be, for example, invalid placement of features within a defined tolerance, duplication of features and invalid overlap of features. |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | Example 1: Towers with identical attribution and within search tolerance (search tolerance = 10 m)  |

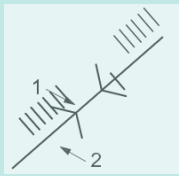
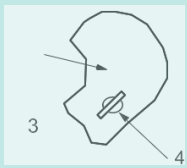
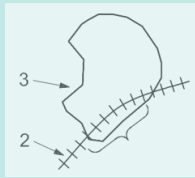
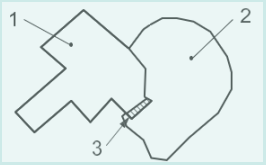
| | | |
|---|---|---|
| | | Example 2: Bridge has invalid Transportation. Use Category of Road<br>  <br><br>Example 3: Invalid placement of Airport inside a Lake<br>  <br><br>Example 4: Invalid overlap of area feature Lake within line feature Railroad<br>  <br><br>Key : 1. Bridge, 2. Railroad, 3. Lake, 4. Airport |
| 13 | Identifier | 4 |

Table 5: Number of Invalid Overlaps of Surfaces

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of invalid overlaps of surfaces |
| 2 | Alias | Overlapping surfaces |
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Conceptual consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Total number of erroneous overlaps within the data |

| | | |
|---|---|---|
| 7 | Description | Which surfaces may overlap and which shall not be application dependent. Not all overlapping surfaces are necessarily erroneous.<br>When reporting this data quality measure, the types of feature classes corresponding to the illegal overlapping surfaces shall be reported as well. |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example |  <br><br>Key<br>1 Surface 1<br>2 Surface 2<br>3 Overlapping Area |
| 13 | Identifier | 5 |

ii. Domain Consistency

The data quality measures for the data quality element domain consistency are provided below.

Table 6: Number of Items Not In Conformance With Their Value Domain

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of items not in conformance with their value domain |
| 2 | Alias | – |

| | | |
|---|---|---|
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Domain consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all items in the dataset that are not in conformance with their value domain |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 6 |

iii. Format Consistency

The data quality measures for the data quality element format consistency are provided below.

Table 7: Physical Structure Conflicts

| Line | Component | Description |
|---|---|---|
| 1 | Name | Physical structure conflicts |
| 2 | Alias | – |
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Format consistency |
| 5 | Data quality basic measure | Error count |

| | | |
|---|---|---|
| 6 | Definition | Count of all items in the dataset that are stored in conflict with the physical structure of the dataset |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 7 |

iv.Topological Consistency

The data quality measures in below are designed to test the topological consistency of geometric representations of features.

Table 8: Connectivity Error

| Line | Component | Description |
|---|---|---|
| 1 | Name | Connectivity Error |
| 2 | Alias | – |
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Topological consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all connectivity errors in the dataset having line features. |
| 7 | Description | – |
| 8 | Parameter | – |

| | | |
|---|---|---|
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 8 |

Table 9: Adjacency Error

| Line | Component | Description |
|---|---|---|
| 1 | Name | Adjacency Error |
| 2 | Alias | – |
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Topological consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all adjacency errors where adjacent features in a polygon layer do not properly share a common boundary. |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 9 |

Table 10: Containment Error

| Line | Component | Description |
|---|---|---|
| 1 | Name | Physical structure conflicts |
| 2 | Alias | – |
| 3 | Data quality element group | Logical consistency |
| 4 | Data quality element | Topological consistency |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Count of all containment errors where one feature or set of features (such as points, lines and polygons) does not completely enclose within a polygon. |
| 7 | Description | – |
| 8 | Parameter | – |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 10 |

### 3. Positional Accuracy

i. Absolute or external accuracy

The data quality measures for positional uncertainty in general of the data quality element absolute provided below.

Table 11: Number of Positional Uncertainties Above a Given Threshold

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of positional uncertainties above a given threshold |
| 2 | Alias | – |
| 3 | Data quality element group | Positional accuracy |
| 4 | Data quality element | Absolute or external accuracy |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Number of positional uncertainties above a given threshold for a set of positions<br>The errors are defined as the distance between a measured position and what is considered as the corresponding true position. |
| 7 | Description | For a number of points *(N)*, the measured positions are given as *xmi, ymi* and *zmi* coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, *xti, yti* and *zti,* are considered to represent the true positions. The calculation of ei is given by the data quality measure “mean value of positional uncertainties” in one, two and three dimensions.<br>All positional uncertainties above a defined threshold emax ( Ɵi > Ɵmax ) are then counted as error.<br>A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result. |

| | | |
|---|---|---|
| 8 | Parameter | Ɛmax is the threshold for accepted positional uncertainties |
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 11 |

**4. Thematic Accuracy**

i. Classification correctness

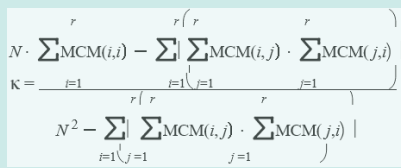
The assignment of an item to a certain class can either be correct or incorrect. Depending on the item that is classified, several data quality measures are given below.

Table 12: Number of Incorrectly Classified Features

| Line | Component | Description |
|---|---|---|
| 1 | Name | Number of incorrectly classified features |
| 2 | Alias | – |
| 3 | Data quality element group | Thematic accuracy |
| 4 | Data quality element | Classification correctness |
| 5 | Data quality basic measure | Error count |
| 6 | Definition | Number of incorrectly classified features |
| 7 | Description | – |
| 8 | Parameter | – |

| Line | Component | Description |
|---|---|---|
| 9 | Data quality value type | Integer |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 12 |

Table 13: Kappa Coefficient

| Line | Component | Description |
|---|---|---|
| 1 | Name | Kappa Coefficient |
| 2 | Alias | – |
| 3 | Data quality element group | Thematic accuracy |
| 4 | Data quality element | Classification correctness |
| 5 | Data quality basic measure | – |
| 6 | Definition | Coefficient to quantify the proportion of agreement of assignments to classes by removing misclassifications |
| 7 | Description | With the elements of the misclassification matrix MCM (i, j) given as data quality measure in Table D.62, the kappa coefficient (κ) can be calculated by  *N* is the number of classified items |

| Line | Component | Description |
|---|---|---|
| 8 | Parameter | Number of classes under consideration *r* |
| 9 | Data quality value type | Real |
| 10 | Data quality value structure | – |
| 11 | Source reference | – |
| 12 | Example | – |
| 13 | Identifier | 13 |

# Annexure II: Example of Data Quality Reporting

The example for data quality reporting for each data quality element is based on the existing Data Product Specification of 1/25000 scale Topographic Base Map of Bhutan.

**Example 1: Commission**

| Data Quality Components | Description |
|---|---|
| Quality Requirement | The number of instances is equal to the numbers of the reference data |
| Data Quality Element Group | Completeness |
| Data Quality Element | Commission |
| Data Quality Scope | Instances of the following class:<br>Boundary_line, Boundary Pillars & Control Point (Ground Control<br>Points and Benchmarks) |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided.<br><br>**Name:** Number of excess items<br>**Definition:** Number of items within the dataset that should not have been in the dataset<br>**Description:** 1. Count the number of data included in the reference data class by class.<br>2. Count the number of instances of the dataset class by class.<br>3. Calculate the difference between the number of reference data and the number of instances class by class using the results of above 1.and 2. The sum of absolute values of the differences is the number of errors.<br>**Data Quality Value Type:** Integer |

| | |
|---|---|
| | **Description:** For all instance pairs of the same class, count the number of instances which cross or overlaps<br>**Data quality value type:** Integer |
| Data Quality Evaluation Procedure | Direct Evaluation method is used and full inspection is done on all instances. |
| Conformance Quality Level | Acceptance: The number of errors is zero (0).<br>Rejection: The number of errors is one (1) or more. |

**Example 2: Format Consistency**

| Data Quality Components | Description |
|---|---|
| Quality Requirement | Well-formed XML document |
| Data Quality Element Group | Logical Consistency |
| Data Quality Element | Format Consistency |
| Data Quality Scope | All dataset of Fundamental Geospatial data (Scale 1/25,000) |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided.<br><br>**Name:** physical structure conflicts<br>**Definition:** count of all items in the dataset that are stored in conflict with the physical structure of the dataset<br>**Description:** Using an inspection program (e.g. XML parser), count the number of places where Well-Formed XML is not correctly used within the dataset.<br>**Data quality value type:** Integer |
| Data Quality Evaluation Procedure | Direct Evaluation method is used for full inspection of all instances. |

| Data Quality Components | Description |
|---|---|
| Conformance Quality Level | Acceptance: The number of places whose feature type does not match the feature type defined by application schema is zero (0).<br>Rejection: The number of places whose feature type does not match the feature type defined by application schema is one (1) or more.<br>The dataset which does not achieve this quality requirement is not considered as a finished product. |
| Data Quality Result | Pass: Yes<br>Specification: DPS for Fundamental Geospatial data (Scale 1/25,000)<br>Explanation: All the data specified in the scope conforms to the format consistency specified in DPS |

**Example 3: Topological Consistency**

| Data Quality Components | Description |
|---|---|
| Quality Requirement | Line feature instances do not cross or overlap with other instances within the same class. |
| Data Quality Element Group | Logical Consistency |
| Data Quality Element | Topological Consistency |
| Data Quality Scope | Relationship between instances within the same class of the following class:<br>Boundary_Line, Road_Line, Ropeway_Line, Hydro_Line |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided.<br>**Name:** number of invalid self-intersect errors<br>**Definition:** count of all items in the data that illegally intersect with themselves. |

| | |
|---|---|
| | **Description:** For all instance pairs of the same class, count the number of instances which cross or overlaps<br>**Data quality value type:** Integer |
| Data Quality Evaluation Procedure | Direct Evaluation method is used and full inspection is done on all instances. |
| Conformance Quality Level | Acceptance: The number of errors is zero (0).<br>Rejection: The number of errors is one (1) or more.<br>(the same class) Line feature instances fulfill the given connection conditions at connecting points with other instances within the same class. |
| Data Quality Result | Number of invalid self-intersect : 0<br>valueUnit: count<br>valueRecordType : Integer |

**Example 4: Absolute or External Positional Accuracy**

| Data Quality Components | Description |
|---|---|
| Quality Requirement | To have the coordinates equal to the coordinate values accepted. |
| Data Quality Element Group | Positional Accuracy |
| Data Quality Element | Absolute or External Positional Accuracy |
| Data Quality Scope | Control_Point |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided.<br>**Name:** number of positional uncertainties above a given threshold<br>**Definition:** number of positional uncertainties above a given threshold for a set of positions. The errors are defined as the distance between a measured position and what is considered as the corresponding true position. |

| | |
|---|---|
| | **Description:** 1. Count the number of all instances of the target feature class.<br>2. Count the number of errors.<br>3. Calculate error ratio of each class based on the results of 1 and 2 above.<br>**Data quality value type:** Integer |
| Data Quality Evaluation Procedure | Direct Evaluation method is used and full inspection is done on all instances. |
| Conformance Quality Level | Acceptance: The number of errors is zero (0).<br>Rejection: The number of errors is one (1) or more.<br>The standard deviation of the horizontal position errors calculated by check survey is less than acceptable quality level. |
| Data Quality Result | number of positional uncertainties above a given threshold : 0<br>valueUnit: count<br>valueRecordType : Integer |

**Example 5: Classification Correctness**

| Data Quality Components | Description |
|---|---|
| Quality Requirement | Classification code, feature type and visibility flag are set correctly. |
| Data Quality Element Group | Thematic Accuracy |
| Data Quality Element | Classification Correctness |
| Data Quality Scope | The instances which are given incorrect classification code, feature type or visibility flag are treated as error instances. |
| Data Quality Measure | A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided. |

| | |
|---|---|
| | **Name:** number of incorrectly classified features<br>**Definition:** number of incorrectly classified features<br>**Description:** Count the number of error instances<br>**Data quality value type:** Integer |
| Data Quality Evaluation Procedure | Direct Evaluation method based on sampling inspection is done. |
| Conformance Quality Level | Acceptance: The number of errors is zero (0).<br>Rejection: The number of errors is one (1) or more. |
| Data Quality Result | number of incorrectly classified features : 0<br>valueUnit: count<br>valueRecordType : Integer |

National Spatial Data Infrastructure
Centre for Geo-Information