GOVERNMENT GAZETTE

OF THE

REPUBLIC OF NAMIBIA

N$7.20 WINDHOEK - 7 October 2016 No. 6145

CONTENTS

Page

General Notice

NAMIBIA STATISTICS AGENCY

No. 406 2016

DATA QUALITY STANDARD FOR THE PURCHASE, CAPTURE, COLLECTION, PRODUCTION AND DISSEMINATION OF GEOSPATIAL DATA IN NAMIBIA

TABLE OF CONTENTS

1. INTRODUCTION

The term “data quality” is complex and has multiple viewpoints. Data quality is not absolute, but varies from producer to producer since requirements for usage differ. Also data users need a specific view that helps them to evaluate how much the data fulfil their actual requirements. Since

requirements also differ from user to user it is clear that more than one description of the term data quality can be valid. Furthermore, spatial data and information is being produced by different thematic communities that have developed their own terms. In case of geographic information the relevant existing standards like ISO 19113, 19115, 19138 and 19157 of ISO TC 211 provide a good basis for coherent data quality descriptions. In Namibia data quality and related Metadata quality elements should be more stringent for reference data than those on thematic data themes.

Quality information shall be reported as *metadata* and as a *standalone quality report* as per the ISO 19157:2013 documentation and international best practices. These two mechanisms complement each other by allowing the reporting of data quality evaluation with different levels of detail. The metadata aims at providing short, synthetic and generally-structured information to enable metadata interoperability and web services usage. The standalone quality report may be used to provide fully detailed information about the data quality evaluation. The standalone quality report is to be provided and attached to the data set or product for direct assessment.

2. SCOPE

The scope of this standard is to evaluate the quality of geospatial data in Namibia in alignment to international standards and best practices. Elements of the data quality standard for Namibia shall be useful to add to the other relevant spatial data standards. Some of the elements will be less useful depending on the type of geospatial product, e.g. orthoimagery has different quality elements than vector data. The data quality standard has to be used before submitting metadata or actual geospatial data to the National Spatial Data Infrastructure (NSDI) and will be used by the Namibia Statistics Agency (NSA) to assess or verify data quality. Each government body must declare a quality certification of their spatial data and may adopt a self-certification and quality-audit process according to this data quality standard.

The NSA may submit this self-certification and audit result to the Committee for Spatial Data for verification. It is only after certification and audit are approved that such geospatial data shall form a part of the NSDI. Geospatial data and metadata will not be publicised until they have passed the self-quality audit. The NSA can notify the public about the status of a dataset whether such dataset meets its intended purpose or not.

Accuracy standards for cadastral data are prescribed in accordance with the Land Survey Act, 1993 (No. 33 of 1993). Base control surveys should be performed to a positional confidence level consistent with the cadastral, engineering or construction application or standards in those fields. An estimated accuracy statement is applicable to CADD and GIS or databases that may be compiled from a variety of sources containing known or unknown accuracy reliability.

3. CONFORMANCE

A spatial dataset or product claiming conformance to the Namibia NSDI standard on data quality shall fulfil all the requirements described in the data quality elements table (Table 2). The NSA can independently validate the data quality reported by a data custodian or producer. The validation will include checking whether the quality evaluation process included all of the mandatory and conditional quality elements as specified in the dataset's quality evaluation report. Note that ISO 19157-1:2013 evaluation format may be adapted and incorporated in the quality assessment practical hand book to be developed by the NSA.

4. SPATIAL DATA QUALITY ELEMENTS

There are many definitions of data quality as this term is a function of many intangible properties such as completeness, consistency, and accuracy. However, data is generally considered high quality if "they are fit for their intended uses in operations, decision making and planning." Another term that need amplification in this document is 'error.' An error is a discrepancy between the encoded and

actual value of a particular attribute for a given spatial or non-spatial entity. There are many errors of which three broad categories can be discerned. They are positional, thematic and temporal errors. The definition of error herein assumes the existence of some objective, external and unbiased reality against which such errors can be evaluated e.g. conducting ground verification.

Generally and in compliance with ISO standards, data quality of a spatial dataset can be described by two components of data quality elements and data quality overview elements. Data quality elements describe how well a dataset meets the criteria set forth in its product standard and how well it provides quantitative quality information. Data quality overview elements provide general non-quantitative information.

Table 1: Categories of data quality elements

DATA QUALITY ELEMENTS	DATA QUALITY OVERVIEW ELEMENTS
Completeness	Purpose
Logical consistency	Usage
Positional accuracy	Lineage
Temporal quality	
Thematic accuracy	
Quantitative thematic accuracy	

The most common spatial data quality evaluation methods on the elements 'positional and attribute accuracy' are adopted in this standard.

5. ACCURACY AND EVALUATION METHODS

Accuracy is the inverse of error. It refers to whether the data values (spatial or non-spatial) stored for a spatial entity are the correct values. To be correct a data value must be the right value and must be represented in a consistent and unambiguous form. Three (3) elements of accuracy in spatial data are positional, thematic and temporal accuracies.

5.1 Positional Accuracy

Position of features is generally represented by two sets of coordinates; horizontal and vertical coordinates. Positional accuracy refers to the accuracy of the spatial geometric component of the data or database based on dimensionality. The accuracy is reported in metric units / Euclidean distance or the data unit e.g. positional accuracy in meters. In Namibia the metric used is the root mean square error (RMSE) which is computed as the square root of the mean of the squared errors. The horizontal RMSE is a two dimensions giving x and y coordinates. The independent values are obtained from a source of high accuracy while test values are measured.

This standard requires a 95 percent (%) confidence level for whatever positional accuracy requirement on fundamental datasets. To achieve this, the vertical RMSE is multiplied by a factor of 1.96 and the horizontal RMSE is multiplied by 1.7308**.**

5.1.1 The horizontal spatial accuracy in this standard represents the circular error of a data set's horizontal coordinates at a specified confidence level *(95% Accuracy = 1.7308 x RMSE for fundamental datasets).*

5.1.2 Vertical spatial accuracy is defined by the linear error of a data set's vertical coordinates at a specified confidence level *(95% Accuracy = 1.96 x RMSE for fundamental datasets).*

5.1.3 Accuracy reported at the 95% confidence level in this standard shall mean that 95% of positions in the dataset or map must have an error with respect to true ground position that is equal to or smaller than the reported accuracy value.

5.1.4 Notice that the confidence level for the vertical accuracy above is appropriate when tested in non-vegetated areas/ landscapes.

5.1.5 For vegetated landscapes, testing must be based on the 95th percentile confidence interval especially for LIDAR data.

5.1.6 The above confidence levels are minimum requirements for fundamental datasets. The data custodian can define accuracy requirements higher than the defined minimum requirements in this profile.

5.1.7 The reported accuracy value is the cumulative result of all uncertainties. The reference scheme for radial or linear errors must be defined as relative to absolute geospatial reference networks for cadastral applications or a superior source of higher accuracy for all other fundamental datasets.

5.1.8 Spatial data may be compiled to comply with one level of accuracy in the vertical component and another in the horizontal component.

5.1.9 In Namibia the key to certification of spatial data shall depend on *the ability of the geospatial data in meeting the purpose for which such data was collected.* Generally data custodians are responsible for defining the accuracy requirements of their projects.

5.1.10 However, where a product shall be consumed by two or more data custodians, the definition of the required accuracy level shall be done in a collaborative manner when collecting such data within the framework of the NSDI. This is so in order to allow for maximum utilisation of the collected datasets and avoidance of duplication.

5.2 Thematic Accuracy

This defines the closeness of attribute values to their true value. Unlike positional accuracy, thematic accuracy depends on domain of attributes, both qualitative and quantitative attributes of the mapped feature. Metrics of thematic accuracy or attribute accuracy vary with measurement scale. For quantitative attributes, metrics are similar to those used to measure positional accuracy for point features i.e. RMSE. This is so because quantitative attributes are seen as statistical surfaces similar to data such as elevation.

For categorical attributes, this standard requires a classification accuracy assessment to be conducted and a classification error matrix generated. The matrix enables the calculation of user, producer and overall thematic accuracies from ground truth data or data of superior accuracy.

*** *Producer Accuracy* = Ground Truth Data for the class e.g. water / Total Ground Truth Data (rows)

*** *User Accuracy* = Ground Truth Data for the Class / Number of Classified Data for the Class (columns)

*** *Overall accuracy* = Ground Truth Data / Overall Total of Classified Data

5.2.1 Accuracy assessment shall be conducted on the major attributes of a dataset. These are the attributes which when left out will render a dataset incomplete in context, semantic and theme e.g. a soils dataset must have the minimum attributes according to a soil classification scheme.

5.2.2 Quantitative thematic attributes shall be assessed similar to positional accuracy for point features i.e. RMSE.

5.2.3 For categorical attributes a classification accuracy assessment must conducted and a classification error matrix generated.

5.2.4 It is mandatory to report both *producer* and *user thematic accuracies* for categorical attributes. In addition a producer may also report the overall accuracy of a dataset.

5.3 Temporal Quality or Accuracy

This provides the date of data observation, type of update (creation, modification, deletion, unchanged) and validity periods for spatial data. Generally temporal characteristics of spatial objects are handled as special or thematic attributes of the appropriate object type or classes. Some elements of this are conditional as they are valid on dataset where time measurement is present. Temporal accuracy is subdivided into accuracy of a time measurement, temporal consistency and temporal validity.

5.3.1 Where temporal quality is required, it is compulsory to report the *accuracy of a time measurement* which is the closeness of the reported time measurements to values accepted as or known to be true. For example, the correctness of dates of completion of construction of schools in a schools database. Notice that the time and date format in this standard are reported in accordance with ISO 8601 – YYYY-MM-DD.

5.3.2 Where a data or part of it refer to the order of events, evaluate and report *temporal consistency* in the temporal quality element. This refer to the correctness of the order of events in a dataset e.g. START DATE and END DATE. Temporal consistency may also be assessed and reported under conceptual consistency (5.5.1).

5.3.3 It is also required to check the *temporal validity* of data which refer to the validity of data with respect to time. Check if data was captured on the date specified in the lineage. Report the number or percentage of items failing the check.

5.4 Completeness

Completeness describes the exhaustiveness of a set of features and their attributes in spatial data or database. It refers to the commission or omission of an entity or entity attribute in the spatial data as related to their real world entities. The completeness of data or database indicates the level of credibility of how accurate features are represented. As such completeness is dependent on accuracy, both thematic and geographic closeness of the features in the universe of discourse.

5.4.1 It is mandatory to report any errors of commission present in a dataset. The error of commission refers to any excess data present in a dataset.

5.4.2 Producers of spatial data shall also report the existence of any error of omission in a dataset. The error of omission refers to any data absent or missing in a dataset.

5.5 Logical Consistency

This represents the degree of achieved reliability of logical rules and connections in data structures that indicates structural integrity of such data. Logical consistency is further classified into conceptual, domain, format and topological consistencies.

5.5.1 It is mandatory to report the level of *conceptual consistency* of a dataset. This refer to the adherence to rules of the conceptual schema e.g. the name of all classes (feature types, data types, etc.), attribute names, and also domains for all attributes, etc.

5.5.2 Domains of values are usually described by the conceptual schema of the application and can be reported as part of the conceptual consistency in 5.5.1 or separately as *domain consistency*. Where domain consistency, which is the adherence of values to the value domains, exists

on a dataset but was not reported in 5.5.1, then it is required to report this separately under domain consistency.

5.5.3 In accordance with the recommendation in the ISO 19157:2013 document, *format consistency* must mainly be used as the first quality evaluation testing whether the data set is in the correct format according to the product specification. Format consistency is the degree to which data is stored in accordance with the physical structure of the dataset e.g. GML, text, vector (point, line or polygon), raster, etc.

5.5.4 *Topological consistency* refer to the lack of topological errors e.g. unclosed polygons, dangling lines, etc. It shall be mandatory to report any topological errors expected on a dataset as part of a dataset's logical consistency. Note that topological consistency can also be reported as part of conceptual consistency.

5.6 Lineage

Lineage describes the history of a dataset and recounts the life cycle of a data set from collection and acquisition through compilation and derivation to its current form (ISO 19115-1:2014). It describes what happens to data as it goes through diverse processes. It helps provide visibility into the analytics pipeline and simplifies tracing errors back to their sources.

6. ACCURACY TESTING AND VERIFICATION

6.1 Geospatial data and related products that are tested and found to comply with this standard shall have a certification statement. The NSDI secretariat shall also issue a metadata certification for data that meet the required standards for the NSDI.

6.2 Due to the high cost of field testing, not all geospatial data may be tested. In such cases, the accuracy statement shall clearly indicate that the procedural mapping standards were designed and performed to meet a certain accuracy standard (project dependent, National Mapping Accuracy Standard, etc.), but that the accuracy is estimated.

6.3 An estimated accuracy statement is especially applicable to computer-aided design and drafting (CADD) and Geographic Information System (GIS), or databases that may be compiled from a variety of sources containing known or unknown accuracy reliability.

6.4 Report accuracy of spatial data in ground units using either metric units, or units consistent with the measurement units for thematic accuracy.

6.5 If data of varying accuracies are composited and cannot be separately identified and the data set is tested, report the accuracy value for the composited data. If a composited data set is not tested, report the accuracy value for the least accurate data set component.

6.6 A data quality evaluation reporting template shall be developed and provided to spatial data producers by the NSDI secretariat.

6.7 The NSDI secretariat shall strive to build capacity in participating NSDI institutions to ensure that quality assessments are conducted. The secretariat shall develop materials specifically aimed at guiding the implementation of this standard. These will include a spatial data quality assessment practical handbook.

7. POSITIONAL AND THEMATIC ACCURACY STATEMENTS

7.1 Data quality statements are part of a quality evaluation report and also incorporated in the metadata.

7.2 Once positional or thematic accuracy of a data set has been determined, it is important to report that value in a consistent and meaningful manner.

7.3 For Namibia two (2) reporting statements for positional accuracy are for tested and compiled datasets as follows:

(a) Tested _____ (meters, unit) (horizontal, vertical) accuracy at 95% confidence level.

(b) Compiled to meet _____ (meters, unit) (horizontal, vertical) accuracy at 95% confidence level.

7.4 Estimated accuracy shall be categorised under compiled datasets in this standard. Note that estimated accuracy does not apply to cadastral and engineering applications. These applications generally require tested accuracy with clearly defined closing errors.

7.5 As per the requirements in the standard on the manner of capturing of spatial data including any application for exemption from such manner, where GPS technology is used to capture GIS data, framework datasets shall be captured using mapping grade receivers.

8. REFERENCE DATUMS AND COORDINATE SYSTEMS

8.1 Reference datums

8.1.1 Spatial data quality standard requires that features are mapped based on the correct reference datum and coordinate system.

8.1.2 The referencing standard for geospatial data for Namibia as it is prescribed by the surveyor-general in the ministry responsible for lands shall apply in this standard. It is compulsory for all new data acquisition and is applicable for sharing data.

8.1.3 The World Geodetic System 1984 (WGS84) datum shall be used as the horizontal datum for all new digital orthoimagery and all vector data.

8.1.4 For internal viewing, mapping, geo-statistical or other purposes within institutions and for a limited number of end-users requiring special projections, different reference datums and coordinate systems may be used. However when sharing such datasets to a wider audience, conversion to nationally acceptable reference datums and coordinate systems must first be done.

8.2 Coordinate systems

8.2.1 A common method for referencing coordinate positions on the Earth is essential for integrating geospatial data. While it is desirable that fundamental or framework data in Namibia be described by longitude and latitude coordinates, orthoimagery is more often represented in a grid coordinate system such as Universal Transverse Mercator (UTM).

8.2.2 The UTM grid system is accepted as is any well-defined projection system allowing conversion to Namibian projection standards as long as all the conversion parameters are specified. The recommended projection for Namibia is a modified Transverse Mercator as follows:

- Projection Algorithm: Transverse Mercator
- Projection Parameters:
  - Longitude of origin: 17:00:00.000
  - Latitude of origin: 0:00:00.000

- False X: 600000.00
- False Y: 10000000.00
- Scale reduction factor: 1

- Geodetic Datum: WGS84
- Vertical ellipsoid: WGS84
- Vertical Datum Reference: Geoid (orthometric) or the WGS 84 ellipsoid but this must be documented

8.2.3 A second coordinate system that is accepted by this standard is the geographic latitude longitude with datum WGS 84.

8.3 Datum Conversions

8.3.1 Conversions between data using the Schwarzeck datum and the geocentric WGS 84 datum using conventional software may generate considerable errors in the absence of grid based transformation systems.

8.3.2 The vector and imagery data using the Schwarzeck datum should not be transformed when high accuracy is required unless a reliable grid based transformation file is available for Namibia.

8.3.3 The metadata should indicate which conversion method is used in cases where high accuracy is not required.

9. TABLE OF SPATIAL DATA QUALITY ELEMENTS FOR NAMIBIA

9.1 It is required that when collecting, creating, modifying, and sharing spatial datasets in Namibia, the data quality elements in Table 2 be complied to in accordance with the ISO 19157:2013 data quality standard.

9.2 The data quality elements and data quality overview elements must be entered into the associated metadata files of the associated datasets.

9.3 Description must be provided for all the mandatory fields of this standard in the metadata profile (including a quality statement under positional, thematic and temporal accuracy where applicable).

9.4 Conditional data quality elements are mainly those elements which might not be necessary on some datasets, but are required on other datasets. Thus, where a conditional element applies the metadata field shall be mandatory.

TABLE 2: NATIONAL SPATIAL DATA QUALITY ELEMENTS FOR NAMIBIA

CORE DATA QUALITY ELEMENT	DATA TYPE	DESCRIPTION	Mandatory (M), Conditional (C) and Optional (O)
Completeness	Free Text	presence or absence of features, their attributes and relationships	
* *Commission*		excess data present in dataset (sub element of completeness)	M
**Omission*		data absent from a dataset (sub element of completeness)	M
Logical consistency	Free Text	degree of achieved reliability of logical rules and connections in data structures that indicates structural integrity of such data.	
**Conceptual consistency*		adherence to rules of the conceptual scheme (sub element of logical consistency)	O
**Domain consistency*		adherence of values to the value domains (sub element of logical consistency)	O
**Format consistency*		degree to which data is stored in accordance with the physical structure of the dataset (sub element of logical consistency)	M
**Topological consistency*		correctness of the explicitly encoded topological characteristics of a dataset (sub element of logical consistency)	M
Positional accuracy	Free Text	accuracy of the position of features.	
**Absolute accuracy*		mandatory for all vector datasets	M
**Relative accuracy*		optional for all vectors	O
**Gridded raw data position accuracy*		closeness of gridded data position values to values accepted	M
**Gridded raw data spatial resolution*		provide the ground dimensions of the picture elements (pixel) making up the digital image	M
Temporal quality	Free Text	accuracy of the temporal attributes and temporal relationships of features	
**Temporal measurement*		correctness of reported time measurements in the data (provide percentage of correctness)	M
**Temporal consistency*		for ordered events, report percentage of correctness of the order of events in dataset e.g. what occurred first, second, etc.	C
**Temporal validity*		check if data was captured on the date specified in the lineage. Report the number or percentage of items failing the check.	M

CORE DATA QUALITY ELEMENT	DATA TYPE	DESCRIPTION	Mandatory (M), Conditional (C) and Optional (O)
Thematic accuracy	Free Text	accuracy of thematic attributes and correctness of quantitative and non-quantitative attributes and of the classifications of features and their relationships. This includes gridded classified dataset e.g. national land cover maps.	M
**Non-quantitative attribute accuracy*		accuracy of categorical attributes in percentage. Compulsory for main theme attributes	C
**Quantitative attribute accuracy*		accuracy of quantitative attributes in percentage. Compulsory for main theme attributes	C
Purpose	Free Text	information on the reasons for creating the dataset and on the intended use of the dataset	M
Usage	Free Text	information on the kind of application for which the dataset has been used and usage limitations in the dataset	M
Lineage	Free Text	describe the history of the dataset including any analytical processing conducted on the dataset	M

APPENDIX 1: TERMS AND DEFINITIONS BASED ON ISO GLOSSARY OF TERMS

accuracy
closeness of agreement between a test result or measurement result and the true value
In the data quality ISO Standard, the true value can be a reference value that is accepted as true.

conformance
fulfilment of specified requirements

conformance quality level
threshold value or set of threshold values for data quality results used to determine how well a dataset meets the criteria set forth in its data product specification or user requirements

correctness
correspondence with the universe of discourse

data product specification
detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party

dataset
identifiable collection of data. A data set can be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger data set. Theoretically a data set can be as small as a single feature or feature attribute contained within a larger data set. A hardcopy map or chart can be considered a data set.

dataset series
collection of datasets sharing common characteristics

Direct evaluation method
method of evaluating the quality of a dataset based on inspection of the items within the dataset

feature
abstraction of real world phenomena. A feature may occur as a type or an instance. Feature type or feature instance will be used when only one is meant.

feature attribute
characteristic of a feature. A feature attribute has a name, a data type and a value domain associated with it. A feature attribute for a feature instance also has an attribute value taken from the value domain.

feature instance
individual of a given feature type having specified feature attribute values

feature operation
operation that every instance of a feature type may perform

feature type
class of features having common characteristics

geographic data
data with implicit or explicit reference to a location relative to the Earth

indirect evaluation method
method of evaluating the quality of a dataset based on external knowledge. Examples of external knowledge are data set lineage, such as production method or source data.

metadata
Information about a resource [SOURCE: ISO 191151]

meta quality
Information describing the quality of data quality

quality
degree to which a set of inherent characteristics fulfils requirements

universe of discourse
view of the real or hypothetical world that includes everything of interest

_____