XXVII FIG CONGRESS
11–15 SEPTEMBER 2022
Warsaw, Poland

Volunteering for the future – Geospatial excellence for a better living

Presented at the FIG Congress 2022, Warsaw, Poland 11-15 September 2022 in Warsaw

LADE AGENDA

A Practical Approach to Quality Control of Geographic Information

by Frank Haugan,
Helge Onsrud and Oskar Henriksen

ORGANISED BY FIG SGP

PLATINUM SPONSORS esri Trimble.

February 2022

Namibia Spatial Data Quality and Metadata Handbook

Towards official certification of spatial data in Namibia

Namibia Statistics Agency

NSDI Namibia

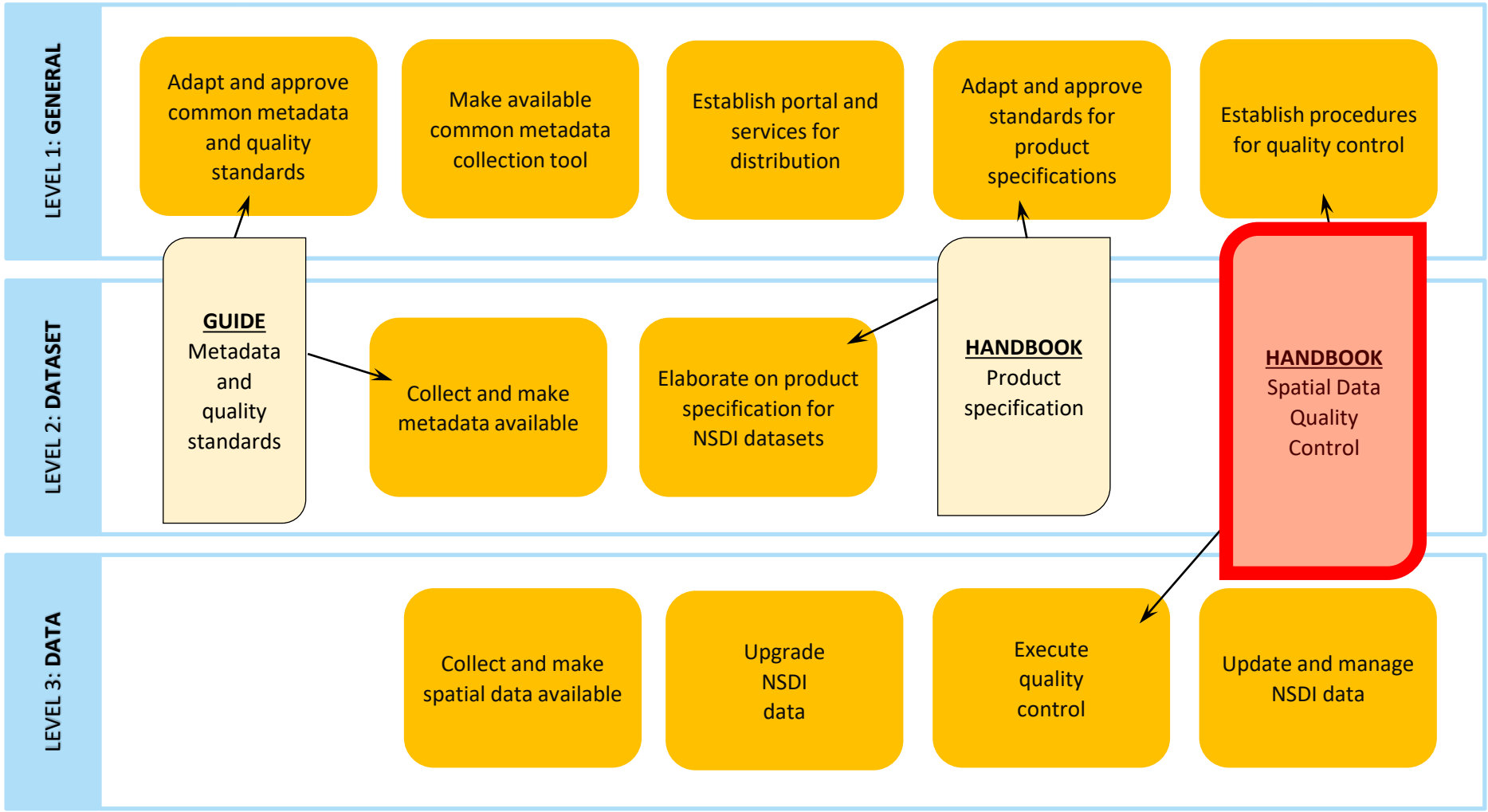
Background and approach

Content of the Handbook

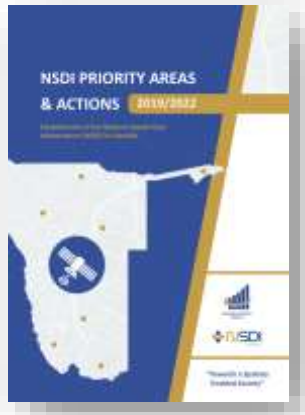
Content of the eLearning platform

The idea is bridging the gap between the complex language in a standard and the spatial data / GIS practitioners

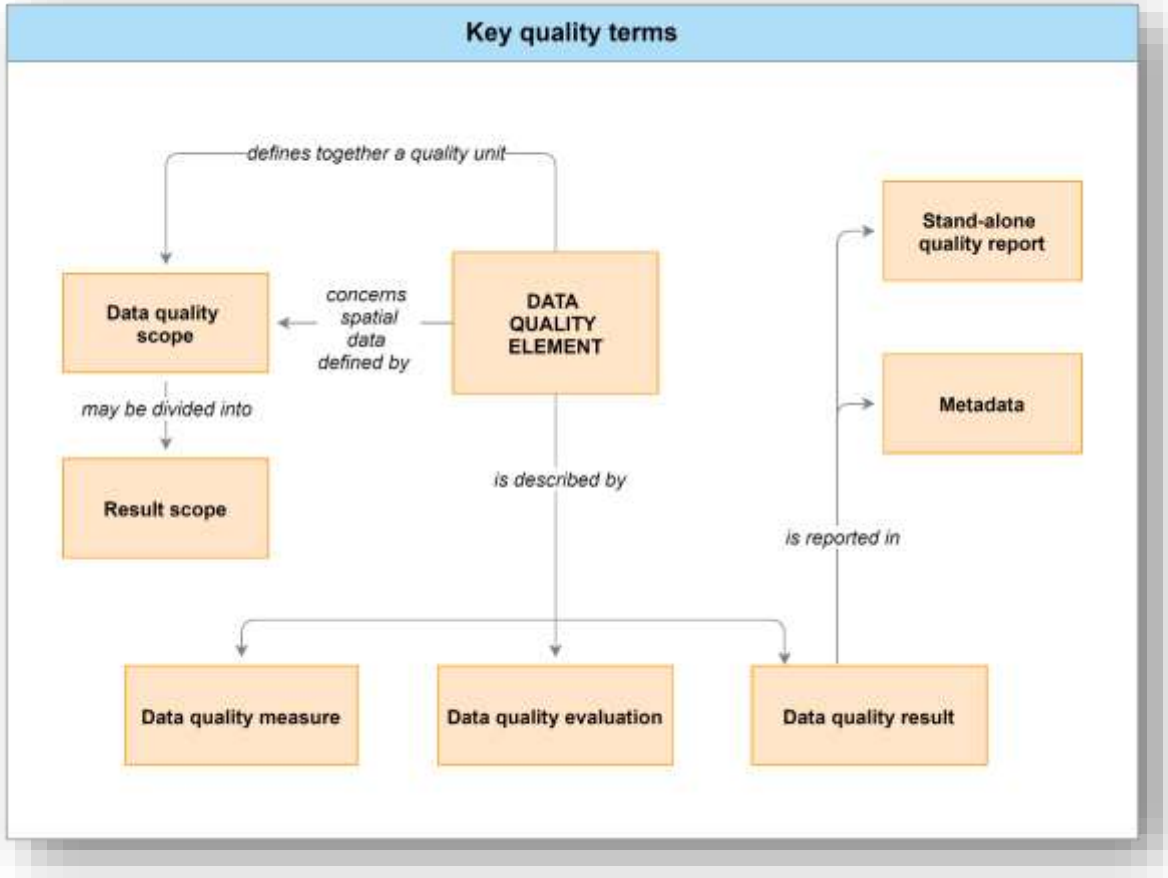
LADE AGENDA


LEVEL 1: GENERAL
Adapt and approve common metadata and quality standards
Make available common metadata collection tool
Establish portal and services for distribution
Adapt and approve standards for product specifications
Establish procedures for quality control
LEVEL 2: DATASET
GUIDE
Metadata and quality standards
Collect and make metadata available
Elaborate on product specification for NSDI datasets
HANDBOOK
Product specification
HANDBOOK
Spatial Data Quality Control
LEVEL 3: DATA
Collect and make spatial data available
Upgrade NSDI data
Execute quality control
Update and manage NSDI data

INTERNATIONAL STANDARD
ISO 19157
Geographic information — Data quality
AMENDMENT 1: Describing data quality using coverages


NSDI PRIORITY AREAS & ACTIONS 2019/2022

GOVERNMENT GAZETTE
OF THE
REPUBLIC OF NAMIBIA
General Notice
Gazette # 6145 of 2016 - Spatial Data Quality Standard
Gazette # 6144 of 2016 - Metadata Standard
Gazette # 6143 of 2016 - Manner and Specification
Gazette # 5684 of 2015 - NSDI Policy


Key quality terms
defines together a quality unit
Data quality scope
concerns spatial data defined by
DATA QUALITY ELEMENT
Stand-alone quality report
Metadata
may be divided into
Result scope
is described by
is reported in
Data quality measure
Data quality evaluation
Data quality result

Content of the Practical Handbook

Generic content

Chapter 1 – Overview and principles
- Purpose and use-cases
- Key quality terms and definitions

Chapter 2 – The process of quality assessment
- Establishing quality requirements
- Specifying quality assessment
- Executing quality assessment

Chapter 3 – Logical consistency
- *Establishing* requirements, included an example of product specification and conformance level
- *Specifying* quality elements for control, quality measures, scope and quality units, evaluation methods, and example of quality assessment
- *Executing* the quality control, included example of quality report and how to update metadata

Chapter 4 – Completeness

Chapter 5 – Thematic accuracy

Chapter 6 – Positional accuracy

Chapter 7 – Temporal quality

Chapter 8 – Quality usability elements

Chapter 9 – Quality reports
- Stand-alone reports
- Reporting quality control with metadata

Chapter 10 – Metadata elements described
- Metadata on data quality
- NSDI metadata quality – UML model
- NSDI metadata quality – Data dictionary

Appendix A – Overall Quality Assessment Checklist

Appendix B – Complete Example of Specifications and Quality Report

Appendix C – Catalogue of Data Quality Measures

Appendix D – Catalogue of Data Quality Evaluation Methods

Appendix E – Sampling and Correction of Quality Measure Accuracy

Examples

Various examples
- All chapters have examples using well known spatial data and GIS knowledge

Road network as example
- The handbook make use of road network to visualise and explain the process of quality control, for example:
  - Preliminary product specification
  - Defining quality measures
  - Define quality units
  - Quality assessment specification
  - Quality report

eLearning material

Interactive web-based eLearning program
- Introduction
- Food for thought
- Key terms
- Questions and answers
- Quiz
- Pitfalls
- Access to handbook and glossary

Content of the Practical Handbook

Use **process diagrams**

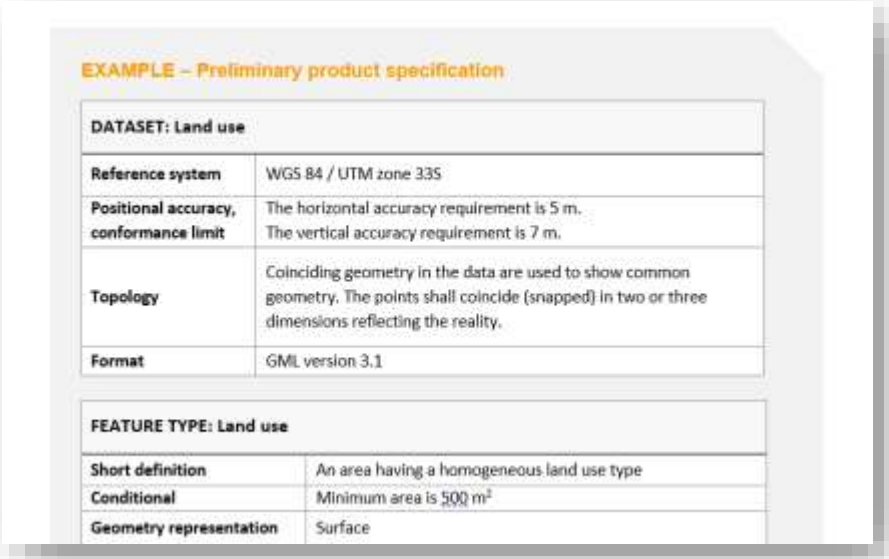
Use **tip frames**

Standard and preliminary product specifications

The handbook refers two types of product specifications. The **standard product specification** is understood as a specification which is according to the ISO 19131:2007 - Data product specifications.

The **preliminary product specification** is understood as spatial data requirements needed to execute quality control.

Use **example frames**



EXAMPLE – Preliminary product specification

DATASET: Land use	
Reference system	WGS 84 / UTM zone 33S
Positional accuracy, conformance limit	The horizontal accuracy requirement is 5 m. The vertical accuracy requirement is 7 m.
Topology	Coinciding geometry in the data are used to show common geometry. The points shall coincide (snapped) in two or three dimensions reflecting the reality.
Format	GML version 3.1

FEATURE TYPE: Land use	
Short definition	An area having a homogeneous land use type
Conditional	Minimum area is 500 m^2
Geometry representation	Surface

Content of the Practical Handbook

Appendix C – Catalogue of Data Quality Measures

QM ID	Quality element	QM-name	QM-definition	QM-description	Value type	Evaluation method (QE ID)
QM1	Logical consistency | format consistency	Format conflict rate	Rate of error in format.	Number of items in the dataset that are stored in conflict with the format specification divided by the total number of items.	Percentage	QE1
QM2	Logical consistency | domain consistency	Domain non-conformance rate	Rate of non-conformance with the value domain. (Feature names, attribute names, code lists etc.)	Number of items in the dataset that are not in conformance with their value domain divided by the total number of items.	Percentage	QE1
QM3	Logical consistency | conceptual consistency	Count of non-compliance cases with conceptual schema	Count of non-conformance with conceptual schema. (Rules to ensure that real-world features are represented correctly.)	Number of items in the dataset not compliant with the conceptual schema.	Integer	QE8
QM4	Logical consistency | topological consistency	Count of non-conformance coinciding lines and points	Count of curves and points that logically coincide in the real world, but not in the dataset.	Number of curves and points that should be coinciding, but do not coincide.	Integer	QE8
QM5	Logical consistency | topological consistency	Count of missing connections overshoot/undershoot	Number of curves that mis-matched due to undershoot or overshoot.	Number of mismatches in the dataset.	Integer	QE8
QM6	Logical consistency | topological consistency	Count of faulty point-curve connections	Count of point curve connections in conflict with the nature of the real-world connection.	Number of conflicts in the dataset.	Integer	QE8
QM7	Logical consistency | topological consistency	Count of invalid self-overlap errors	Count of items that illegally intersect or overlap with themselves.	Number of illegal intersects/ overlaps.	Integer	QE8

Content of the Practical Handbook

Appendix D – Catalogue of Data Quality Evaluation Methods

QE-ID	QE-name	QE-type	QE-description	Remarks
QE1	Rate of occurrence	Different QE types	Calculated as percentage of occurrences of the total number within the scope.	QE type depending on quality unit.
QE2	Linear accuracy assessment	Direct external	(1) The difference between the dataset value, and the real-world value are calculated for each item. (2) The RMSE is calculated from the differences. (3) Gross errors are identified and removed. (4) RMSE is recalculated without gross errors. (5) 95% Confidence level values are calculated by multiplying with 1,96	Values larger than 3 x RMSE (99,73 confidence level) are considered gross errors.
QE3	Two-dimensional positional accuracy assessment	Direct external	(1) The distances between the dataset positions (XY), and the real-world positions are calculated for each point. (2) The RMSE is calculated from the distances. (3) Gross errors are identified and removed. (4) RMSE is recalculated without gross errors. (5) 95% Confidence level values are calculated by multiplying with 1,7308	XY-errors larger than 3 x RMSE (99,73 confidence level) are considered gross errors.
QE4	Point-Line positional accuracy assessment	Direct external	(1) The shortest distance from the measured point to the dataset line is calculated (2) The RMSE is calculated from the distances. (3) The Linear accuracy (QM2) is calculated	See QE2

Content of the Practical Handbook

Appendix B – Complete example of specifications and quality report

a) Example of a **preliminary product specification** for road network

b) Example of a **quality assessment specification** for road network

c) Example of a **quality reports** for road network:
- Identification
- Overview results of quality assessment
- Statement of compliance

Quality unit 1: Format consistency - Road Network - Format conflict rate	
Quality elements	Logical consistency | Format consistency
Quality measure ID and name	QM1: Format conflict rate
Scope	Dataset: Road Network Full inspection
Conformance Level	All data shall adhere to the format rules.
Quality evaluation method ID and name	QE1: Rate of occurrence
Method type (QE-type)	Direct internal
Quality unit 2: Topological consistency - Road Network, RoadLink to RoadNode - Count of non-conformance coinciding lines and points	
Quality elements	Logical consistency | Topological consistency
Quality measure ID and name	QM4: Count of non-conformance coinciding lines and points
Scope	Relation: RoadLink to RoadNode Full inspection
Conformance Level	All road links shall start and end in a road node
Quality evaluation method ID and name	QE8: Count of occurrence
Method type (QE-type)	Direct internal

…. QU3, QU4, QU5… QU20

Content of the Practical Handbook

Appendix B – Complete Example of Specifications and Quality Report

a) Example of a **preliminary product specification** for road network

b) Example of a **quality assessment specification** for road network

c) Example of a **quality reports** for road network:
- Identification
- Overview results quality assessment
- Detailed results quality assessment
- Statement of compliance

Quality unit 14: Absolute accuracy - Road Network, RoadNode, nodePosition (Z) – Linear absolute accuracy	
Quality element	Positional accuracy | Absolute
Quality measure	QM14: Linear absolute accuracy
Feature / Attribute	RoadNode / nodePosition
Population	22 205
Sample	100
Measured RMSE	± 3,25 m.
Correction for sampling	1,12
Correction for control accuracy	1
RMSE for evaluation	± 2,90 m.
RMSE conformance limit	± 2,55 m.
95% confidence level for evaluation	± 5,68 m.
95 % conformance limit (AQL)	± 5,00 m.
Result	Fail
Remark	None

1 of 20 detailed results

Content of the eLearning platform

Generic content

Chapter 1 – Overview and principles
- Purpose and use-cases
- Key quality terms and definitions

Chapter 2 – The process of quality assessment
- Establishing quality requirements
- Specifying quality assessment
- Executing quality assessment

Chapter 3 – Logical consistency
- *Establishing* requirements, included an example of product specification and conformance level
- *Specifying* quality elements for control, quality measures, scope and quality units, evaluation methods, and example of quality assessment
- *Executing* the quality control, included example of quality report and how to update metadata

Chapter 4 – Completeness

Chapter 5 – Thematic accuracy

Chapter 6 – Positional accuracy

Chapter 7 – Temporal quality

Chapter 8 – Quality usability elements

Chapter 9 – Quality reports
- Stand-alone reports
- Reporting quality control with metadata

Chapter 10 – Metadata elements described
- Metadata on data quality
- NSDI metadata quality – UML model
- NSDI metadata quality – Data dictionary

Appendix A – Overall Quality Assessment Checklist

Appendix B – Complete Example of Specifications and Quality Report

Appendix C – Catalogue of Data Quality Measures

Appendix D – Catalogue of Data Quality Evaluation Methods

Appendix E – Sampling and Correction of Quality Measure Accuracy

Examples

Various examples
- All chapters have examples using well known spatial data and GIS knowledge

Road network as example
- The handbook make use of road network to visualise and explain the process of quality control, for example:
  - Preliminary product specification
  - Defining quality measures
  - Define quality units
  - Quality assessment specification
  - Quality report

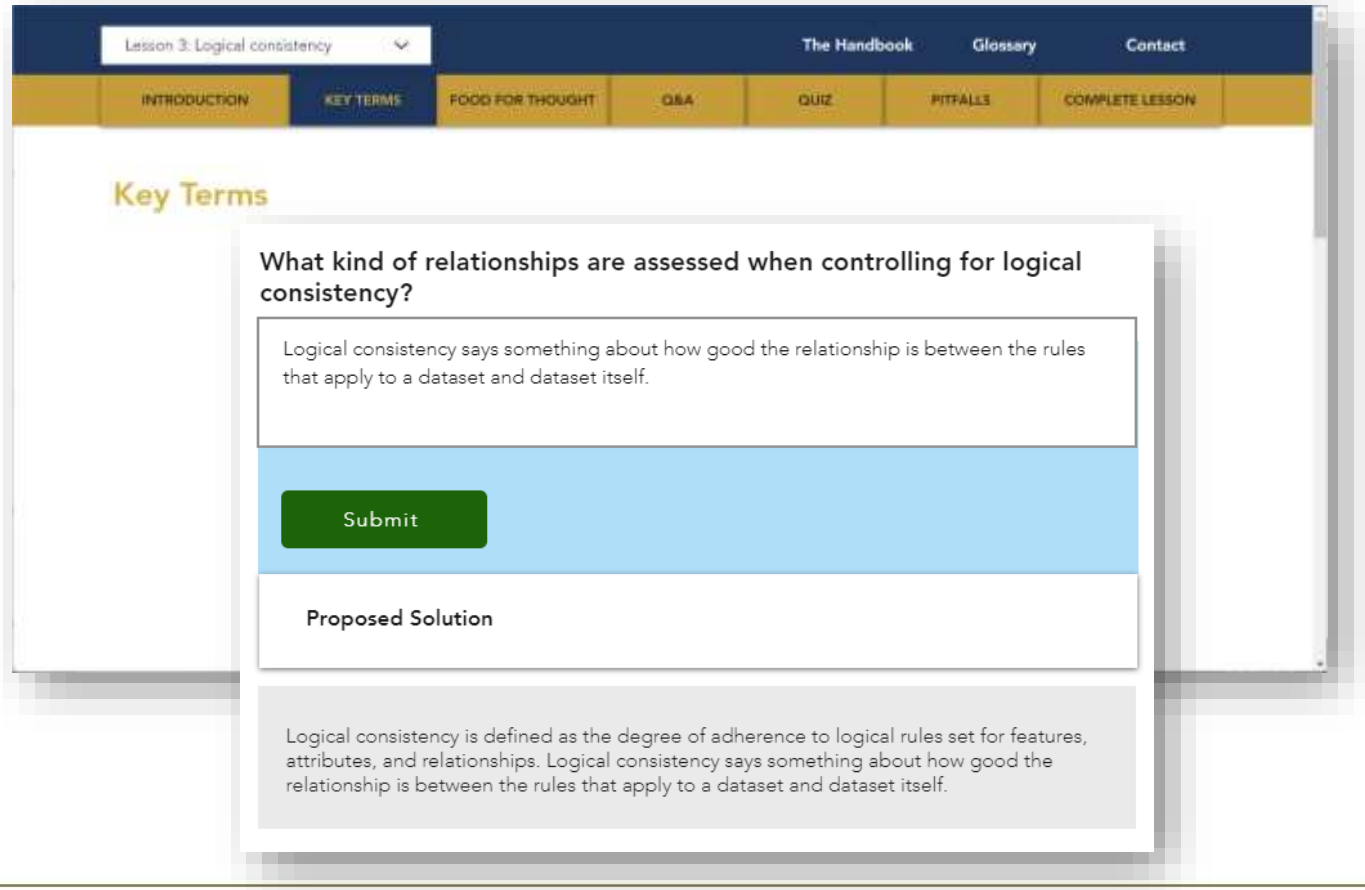
eLearning material

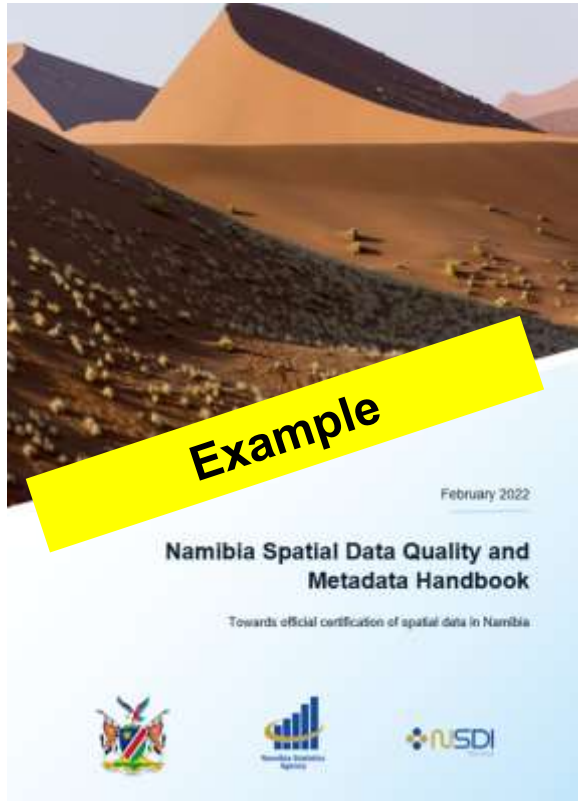
Interactive web-based eLearning program
- Introduction
- Food for thought
- Key terms
- Questions and answers
- Quiz
- Pitfalls
- Access to handbook and glossary

Content of the eLearning platform

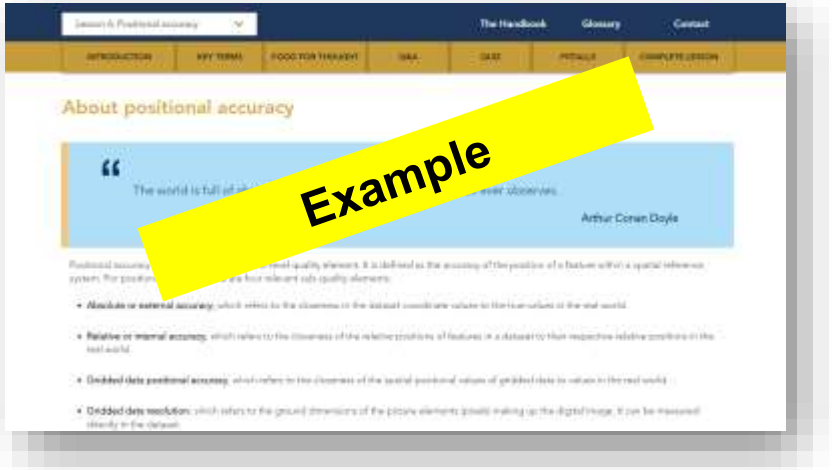
Interactive web-based eLearning program

- Introduction
- Key terms
- Food for thought
- Questions and answers
- Quiz
- Pitfalls





+



+

=

Combine these three sources of information and you will have **valuable insight to spatial data quality and quality control** relevant to needs of the entity and the approval of SDI data.

THANK YOU! QUESTIONS?

CONTACT DETAILS:

Frank Haugan, Director at Lade Agenda Ltd.

frank.haugan@ladeagenda.no

+47 97504773