ISO/TC 211 N 1398

2003-01-29

Number of pages: 72

ISO/TC 211
Geographic information/Geomatics

ISO reference number: 19114

Title: **Text of 19114 Geographic information - Quality evaluation procedures, as sent to the ISO Central Secretariat for publication**

Source: ISO/TC 211 Secretariat

Expected action: For information.

Type of document: Text for publication

Note: This draft will be checked by ISO before being issued as IS

Hyperlink: http://www.isotc211.org/protdoc/211n1398/

Reference: N 1397

ISO/TC 211 Secretariat

Norwegian Technology Centre
Oscars gate 20
P.O. Box 7072 Majorstuen
NO-0306 Oslo, Norway

Telephone: + 47 22 59 01 16
Telefax: + 47 22 59 01 29

E-mail: bjs@nts.no

URL: http://www.isotc211.org/

ISO TC 211/SC

Date: 2003-01-19

ISO 19114:2003(E)

ISO TC 211/SC /WG 3

Secretariat: NSF

Geographic information — Quality evaluation procedures

Information géographique — Procédures d'évaluation de la qualité

Document type: International Standard
Document subtype:
Document stage: (60) Publication
Document language: E

C:\Documents and Settings\bsateroy\Desktop\ISO 19114\FDIS 19114\ISO_FDIS_19114_(E)_Rev.doc STD Version 2.1

Copyright notice

This ISO document is a Draft International Standard and is copyright-protected by ISO. Except as permitted under the applicable laws of the user's country, neither this ISO draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured.

Requests for permission to reproduce should be addressed to either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Reproduction may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

 © ISO 2003 — All rights reserved

Contents

Page

© ISO 2003 — All rights reserved

© ISO 2003 — All rights reserved

© ISO 2003 — All rights reserved

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 19114 was prepared by Technical Committee ISO/TC 211, *Geographic information/Geomatics*.

© ISO 2003 — All rights reserved

Introduction

For the purpose of evaluating the quality of a dataset, clearly defined procedures must be used in a consistent manner. This enables data producers to express how well their product meets the criteria set forth in its product specification and data users to establish the extent to which a dataset meets their requirements. The quality of a dataset is described using two components, a quantitative component and a non-quantitative component. The objective of this document is to provide guidelines for evaluation procedures of quantitative quality information for geographic data in accordance with the quality principles described by ISO 19113. It also offers guidance on reporting quality information.

This International Standard recognizes that a data producer and a data user may view data quality from different perspectives. Conformance quality levels can be set using the data producer's product specification or a data user's data quality requirements. If the data user requires more data quality information than that provided by the data producer, the data user may follow the data producer's data quality evaluation process flow to get the additional information. In this case the data user requirements are treated as a product specification for the purpose of using the data producer process flow.

The quality evaluation procedures described in this International Standard, when applied in accordance with ISO 19113, provide a consistent and standard manner to determine and report a dataset's quality information.

© ISO 2003 — All rights reserved

Geographic information — Quality evaluation procedures

1 Scope

This International Standard provides a framework of procedures for determining and evaluating quality that is applicable to digital geographic datasets, consistent with the data quality principles defined in ISO 19113. This International Standard also establishes a framework for evaluating and reporting data quality results either as part of data quality metadata only or also as a quality evaluation report.

This International Standard is applicable to data producers when providing quality information on how well a dataset conforms to the product specification and to data users attempting to determine whether or not the dataset contains data of sufficient quality to be fit for use in their particular applications.

Although this International Standard is applicable to all types of digital geographic data, its principles can be extended to many other forms of geographic data such as maps, charts and textual documents.

2 Conformance

This International Standard defines three classes of conformance: one for quality evaluation procedures, one for evaluating data quality and one for reporting quality information. The abstract test suites for the three classes of conformance are given in Annex A.

3 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 19113:2002, *Geographic information — Quality principles*

ISO 19115:—[1)] , *Geographic information — Metadata*

4 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 19113 and ISO 19115 (some of which are repeated for convenience) and the following apply.

4,1
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 product specification or user requirements

1) To be published.

© ISO 2003 — All rights reserved

4.2
dataset
identifiable collection of data [ISO 19115]

NOTE A dataset may 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 dataset. For purposes of data quality evaluation, a dataset may be as small as a single feature or feature attribute contained within a larger dataset.

4.3
dataset series
collection of datasets sharing the same product specification [ISO 19115]

4.4
direct evaluation method
method of evaluating the quality of a dataset based on inspection of the items within the dataset

4.5
full inspection
inspection of every item in a dataset [ISO 3534-2]

NOTE Full inspection is also known as 100% inspection.

4.6
indirect evaluation method
method of evaluating the quality of a dataset based on external knowledge

NOTE Examples of external knowledge are dataset lineage, such as production method or source data.

4.7
item
that which can be individually described or considered [ISO 3951]

NOTE An item can be any part of a dataset, such as a feature, feature relationship, feature attribute, or combination of these.

4.8
population
totality of items under consideration [ISO 3534-2]

EXAMPLE 1 All points in a dataset.

EXAMPLE 2 Names of all roads in a certain geographic area.

4.9
reference data
data accepted as representing the universe of discourse, to be used as reference for direct external quality evaluation methods

5 Abbreviated terms

ADQR aggregated data quality results

AQL acceptable quality level [ISO 2859]

RMSE root mean square error

© ISO 2003 — All rights reserved

6 The process for evaluating data quality

6.1 Introduction

A quality evaluation process may be used in different phases of a product life cycle, having different objectives in each phase. The phases of the life cycle considered here are specification, production, delivery, use and update. Annex B describes some specific dataset-related operations to which quality evaluation procedures are applicable.

The process for evaluating data quality is a sequence of steps to produce and report a data quality result. A quality evaluation process consists of the application of quality evaluation procedures to specific dataset-related operations performed by the dataset producer and the dataset user.

Processes for evaluating data quality are applicable to static datasets and to dynamic datasets. Dynamic datasets are datasets that receive updates so frequently that for all practical purposes they are continuously changing. Annex C describes the application of the process to evaluate data quality to dynamic datasets.

6.2 Components of the process

6.2.1 The process flow

The quality evaluation process is a sequence of steps taken to produce a quality evaluation result. Figure 1 illustrates the process flow for evaluating and reporting data quality results.

© ISO 2003 — All rights reserved

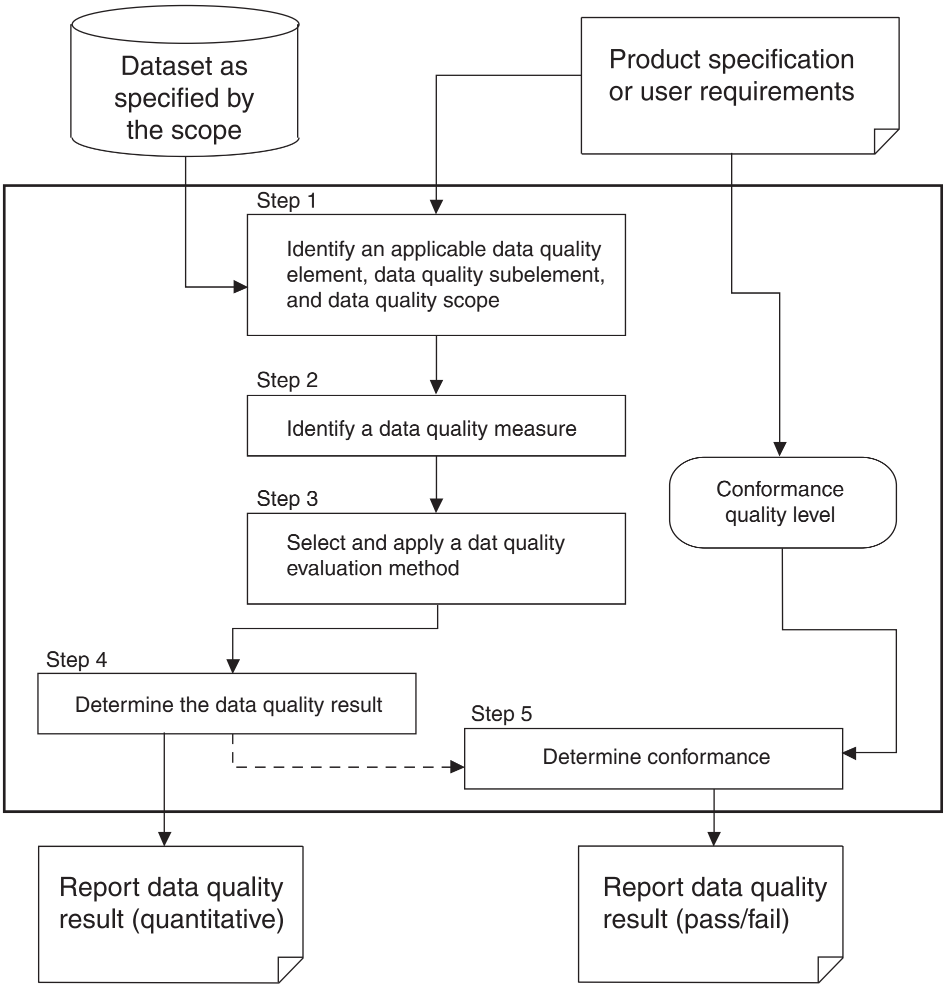


Figure 1 — Evaluating and reporting data quality results

6.2.2 Process steps

Table 1 specifies the process steps.

© ISO 2003 — All rights reserved

Table 1 — Process steps

Process step	Action	Description
1	Identify an applicable data quality element, data quality subelement, and data quality scope	The data quality element, data quality subelement, and data quality scope to be tested shall be identified in accordance with the requirements of ISO 19113. This is repeated for as many different tests as required by the product specification or user requirements.
2	Identify a data quality measure	A data quality measure, data quality value type and, if applicable, a data quality value unit shall be identified for each test to be performed. Annex D provides examples of data quality measures for the data quality elements and data quality subelements given in ISO 19113. Annex D, by these examples, provides assistance to the user in selection of a measure.
3	Select and apply a data quality evaluation method	A data quality evaluation method for each identified data quality measure shall be selected. NOTE A spatial description of the results, e.g., achievable by spatial interpolation of the results, graphical portrayal, etc.) might be useful, corresponding not to a result, but to a different, although related, dataset.
4	Determine the data quality result	A quantitative data quality result, a data quality value or set of data quality values, a data quality value unit and a date is the output of applying the method.
5	Determine conformance	Whenever a conformance quality level has been specified in the product specification or user requirements, the data quality result is compared with it to determine conformance. A conformance data quality result (pass-fail) is the comparison of the quantitative data quality result with a conformance quality level.

7 Data quality evaluation methods

7.1 Classification of data quality evaluation methods

A data quality evaluation procedure is accomplished through the application of one or more data quality evaluation methods. Data quality evaluation methods are divided into two main classes, direct and indirect. Direct methods determine data quality through the comparison of the data with internal and/or external reference information. Indirect methods infer or estimate data quality using information on the data such as lineage. The direct evaluation methods are further subclassified by the source of the information needed to perform the evaluation. Figure 2 depicts this classification structure.

© ISO 2003 — All rights reserved

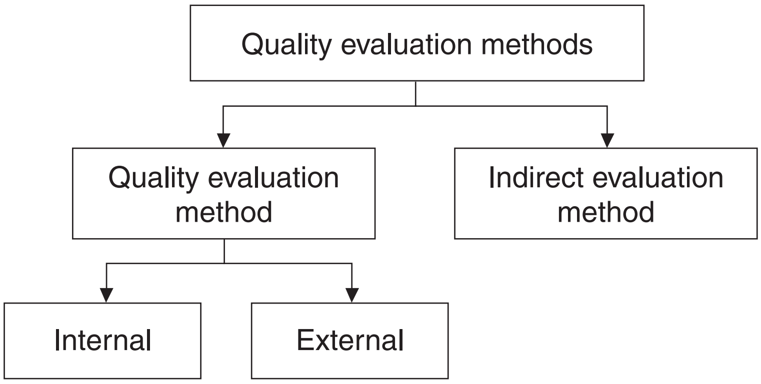


Figure 2 — Classification of data quality evaluation methods (informative)

7.2 Direct evaluation methods

7.2.1 Types of direct evaluation methods

The direct evaluation method is further subdivided into internal and external. All the data needed to perform an internal direct data quality evaluation method is internal to the dataset being evaluated.

EXAMPLE 1 All the data necessary to perform a logical consistency test for topological consistency of boundary closure resides in a topologically structured dataset.

External direct quality evaluation requires reference data external to the dataset being tested.

EXAMPLE 2 The data needed to perform a completeness test for the road names in a dataset requires another information source of road names.

EXAMPLE 3 A positional accuracy test requires a reference dataset or a new survey.

7.2.2 Means of accomplishing direct evaluation

For both external and internal evaluation methods, there are two considerations, automated or non-automated and full inspection or sampling.

NOTE Data quality elements and data quality subelements which are easily checked by automated means include,

logical consistency: format consistency,	[Example: Check data fields for positive entry]
logical consistency: topological consistency,	[Example: Polygon closure]
logical consistency: domain consistency,	[Example: Bounds violations, specified domain value violations]
completeness: omission,	[Example: Comparison check of street names from another file]
completeness: commission,	[Example: Comparison check of street names from another file]
temporal accuracy: temporal consistency.	[Example: Check all records for appropriate range of dates]

7.2.3 Full inspection

Full inspection requires testing every item in the population specified by the data quality scope. Table 2 describes the procedure for full inspection that shall be used.

 © ISO 2003 — All rights reserved

Table 2 — The procedure for full inspection

Procedure step	Description
Define items	An item is a minimum unit to be inspected. An item can be a feature, a feature attribute or a feature relationship.
Inspect items in the data quality scope	Inspect every item it the data quality scope.

NOTE Full inspection is most appropriate for small populations or for tests that can be accomplished by automated means.

7.2.4 Sampling

Sampling requires testing sufficient items in the population in order to achieve a data quality result. Table 3 describes the sampling procedure that shall be used.

Table 3 — Sampling procedure

Procedure step	Description
Define a sampling method	Examples of sampling methods are given in Annex E. Those methods include simple random sampling, stratified sampling (e.g. guided by feature type, a feature relationship or an area), multistage sampling and non-random sampling.
Define items	Item is a minimum unit to be inspected. An item can be a feature, a feature attribute or a feature relationship.
Divide data quality scope (population) into lots	A lot is a collection of items in the data quality scope from which a sample shall be drawn and inspected. Each lot shall, as far as possible, consists of items produced under the same conditions and at the same time.
Divide lots to sampling units	Sampling unit is the area of the lot where inspection is conducted.
Define the sampling ratio or sample size	A sampling ratio gives information on how many items on average are extracted for inspection from each lot.
Select sampling units	Select required number of sampling units so that sampling ratio or sample size for items is fulfilled.
Inspect items in the sampling units	Inspect every item it the sampling units.

The sampling procedure shall be reported in accordance with Clause 8 of this International Standard.

ISO 2859 and ISO 3951 may be applied to sampling for evaluating conformance to a product specification. These standards were originally developed for non-spatial use. Annex E of this International Standard gives examples describing how to apply ISO 2859 and ISO 3951 and provides guidelines on how to define samples and devise sampling methods taking the geographic nature of the data into account.

The reliability of the data quality result should be analysed when using sampling; especially, when using small sample sizes and methods other than simple random sampling.

7.3 Indirect evaluation method

The indirect evaluation method is a method of evaluating the quality of a dataset based on external knowledge. This external knowledge may include, but is not limited to, data quality overview elements and other quality reports on the dataset or data used to produce the dataset.

NOTE 1 This method is recommended only if direct evaluation methods cannot be used.

© ISO 2003 — All rights reserved

NOTE 2 Usage information records uses of a dataset. This is helpful when searching for datasets that have been produced or used for specific purposes.

NOTE 3 Lineage information records information about the production and history of the dataset. It includes information about, for example, source materials to produce a dataset or the production processes applied. This is useful when determining the suitability of a dataset for a given use. An example is lineage metadata relating to a digital terrain model file that has been created by means of stereo-correlation from images captured under certain conditions. Experience tells the evaluator that the horizontal positional RMSE is 10 metres for this type of imagery. Or for example, lineage metadata of a digitised 1:25 000 scale topographic map indicates conformance to a town planner's requirements for a base map.

NOTE 4 Purpose information describes the purpose for which the dataset was produced. A purpose may be in support of a specific requirement or the dataset may be a general purpose dataset for several uses. This is useful when identifying the possible value of a dataset.

7.4 Data quality evaluation examples

Examples of typical methods used and how they may be applied are described in Annexes F, G and H.

8 Reporting data quality evaluation information

8.1 Reporting as metadata

Quantitative quality results shall be reported as metadata in compliance with ISO 19115, which contains the related model and data dictionary.

8.2 Reporting in a quality evaluation report

There are two conditions under which a quality evaluation report shall be produced: (1) When data quality results reported as metadata are only reported as pass/fail, (2) When aggregated data quality results are generated. The report is required in the latter condition to explain how aggregation was done and how to interpret the meaning of the aggregate result. However, a quality evaluation report may be created at any other time, such as to provide more detail than reported as metadata, but a quality evaluation report cannot be used in lieu of reporting as metadata.

A quality evaluation report shall be produced in compliance with Annex I which contains the relevant model and data dictionary.

8.3 Reporting aggregated data quality result

When several quality results are aggregated into a single quality result for reporting the quality of a dataset, the aggregated data quality result shall be reported as metadata and in the data quality report. The data quality result shall be reported as type 'aggregate'. Annex J describes the production of aggregate data quality results and Annex H provides a production example.

© ISO 2003 — All rights reserved

Annex A
(normative)

Abstract test suites

A.1 Introduction

This annex defines three classes of conformance, quality evaluation procedure in A.2, evaluating data quality in A.3 and reporting data quality in A.4. Any quality evaluation procedures claiming conformance with this International Standard shall pass all the requirements described in A.2. Any evaluation of data quality claiming conformance with this International Standard shall pass all the requirements described in A.3. Any report of data quality claiming conformance with this International Standard shall pass all the requirements described in A.4.

NOTE All of the test cases are of test type ‘basic’.

A.2 Abstract test suite for class 1: Quality evaluation procedures

a) Test purpose: To assure the quality evaluation procedure was produced in accordance with this International Standard.

b) Test method: Pass all the requirements described in A.3 and A.4.

c) Reference: A.3 and A.4.

A.3 Abstract test suite for class 2: Evaluating data quality

a) Test purpose: To assure the quality evaluation procedure was produced in accordance with the quality evaluation process in Clause 6.

b) Test method: Compare the quality evaluation procedure with the quality evaluation as appropriate.

c) Reference: ISO 19114, Clause 6.

A.4 Abstract test suite for class 3: Reporting data quality

a) Test purpose: To assure data quality has been reported in accordance with Clause 8.

b) Test method: Compare the quality evaluation reported to assure data quality results were appropriately reported in accordance with Clause 8 and applicable annexes.

c) Reference: ISO 19114, Clause 8.

© ISO 2003 — All rights reserved

Annex B
(informative)

Uses of quality evaluation procedures

B.1 Introduction

Quality evaluation procedures may be used in different phases of a product life cycle. This annex provides examples of stages of a product's life cycle during which quality evaluation procedures may be applied.

B.2 Development of a product specification or user requirements

When developing a product specification or user requirement, quality evaluation procedures may be used to assist in establishing conformance quality levels that should be met by the final product. A product specification or user requirement should include conformance quality levels for the dataset and quality evaluation procedures to be applied during production and updating.

B.3 Quality control during dataset creation

At the production stage, the producer may apply quality evaluation procedures, either explicitly established or not contained in the product specification, as part of the process of quality control. The description of the applied quality evaluation procedures, when used for production quality control, should be reported as lineage metadata, including, but not necessarily limited to, the quality evaluation procedures applied, conformance quality levels established, and the results.

B.4 Inspection for conformance to a product specification

On completion of the production, a quality evaluation process is used to produce and report data quality results. These results may be used to determine whether a dataset conforms to its product specification. If the dataset passes inspection, composed of a set of quality evaluation procedures, the dataset is considered to be ready for use. The results of the inspection operation should be reported in accordance with Clause 8.

The outcome of the inspection will be either acceptance or rejection of the dataset. If the dataset is rejected, then after the data has been corrected, a new inspection will be required before the product can be deemed to be in conformance with the product specification.

B.5 Evaluation of dataset conformance to user requirements

Quality evaluation procedures are used to establish the conformance quality levels for a dataset to meet a user requirement. Indirect as well as direct methods may be used in analyses of dataset conformance to user requirements. The results of the quality evaluation for conformance to user requirements may be reported as usage metadata for the dataset.

B.6 Quality control during dataset update

Quality evaluation procedures are applied to dataset update operations, both to the items being used for update and to benchmark the quality of the dataset after update has occurred. The guidance for use of ISO 19113 and this International Standard on dynamic datasets is given in Annex C.

© ISO 2003 — All rights reserved

Annex C
(informative)

Applying quality evaluation procedures to dynamic datasets

C.1 Introduction

This annex describes how quality evaluation procedures may be applied to dynamic datasets. Here dynamic datasets are defined as datasets that receive updates so frequently that for all practical purposes they are continuously being updated. For example, an online cadastre dataset may receive updates every few minutes. There are basically two ways to determine and report the quality of a dynamic dataset: benchmark and continuous.

C.2 Determining and reporting the quality of a dynamic dataset

C.2.1 Benchmark procedure

The benchmark procedure is based on the establishment of a suitable reporting frequency, for example weekly or tri-monthly, and making a copy of the dataset at the reporting date. Then the copy is tested as if it were a static dataset. This type of testing and reporting will provide quality of the dataset as of the date/time of the copy.

C.2.2 Continuous procedure

The continuous procedure is based on testing the updates and evaluating the impact of the updates. This is equivalent to embedding the quality evaluation procedures given in this International Standard into an ISO 9000-type process-oriented procedure. Since this procedure only can provide current status of the quality of the updated items, it is necessary to combine both benchmark and continuous procedures as described in C.3 in order to establish the quality for the updated database.

C.3 Establishing continuous quality evaluation procedures

C.3.1 Identify the parts

In accordance with the steps given in 6.2 of this International Standard, identify applicable data quality elements and their associated data quality subelements, data quality scopes, data quality measure, and conformance quality levels to be used in the evaluation and reporting of the results.

C.3.2 Select the method to be applied

Select the data quality evaluation method to be applied. Then the evaluation would be on the updated feature and the relationship of that feature with the others within the data quality scope. In a continuous quality evaluation procedures only indirect or internal direct methods may be applied.

EXAMPLES

1) Is the update from a trusted source?

2) Does the update preserve topological consistency?

© ISO 2003 — All rights reserved

3) Does the address of the feature updated retain logical consistency?

C.3.3 Establish a dataset quality reference

Use the benchmark procedure to establish reference values of quality of the dataset for the features and feature attributes within scope to be checked during the continuous testing.

C.3.4 Integrate continuous tests into update process

Integrate the continuous tests into the update process flow so that each proposed update is tested and accepted before it is introduced into the dataset.

C.3.5 Dynamically update data quality results

By integrating the continuous tests into the update process flow, each accepted update causes the current quality results to be adjusted accordingly. This will allow for immediate reports on dataset quality to be generated.

C.4 Periodically re-establish the reference quality of the dataset

All aspects of the quality of a dataset may not be tested through a continuous process-based operation. For example, omission of features may not be found when only updated items are tested. The dataset should be subject to periodic benchmark type quality.

© ISO 2003 — All rights reserved

Annex D
(informative)

Examples of data quality measures

D.1 Introduction

This annex provides simple examples of data quality measures for each data quality element and its associated subelements defined in ISO 19113 to demonstrate how the data quality components relate during an evaluation operation. More detailed examples may be found in other annexes of this International Standard.

For each data quality element and subelement combination, an example data quality scope is given along with example dataset parameters. Then three data quality measures are shown, each designed to demonstrate a different way to evaluate quality. So the examples will be as complete as possible, a data quality date and conformance quality level are given. Finally, an interpretation of the data quality result is given as an example quality result meaning.

While the examples given in this annex are simple, it may be desirable to refer to them in profiles or other documents. Therefore, this annex has a data quality measure identification code which relates the example to the data quality element and data quality subelement.

D.2 Relationship of the data quality components

Table D.1 gives the relationship of the data quality components.

In order to save space, each data quality component has been given a short name that will be used throughout this annex.

© ISO 2003 — All rights reserved

Table D.1 — Relationship of data quality components

Data quality components	Short name[a]	Component domain	Example
Data quality scope	DQ_Scope	Free text	All items classified as houses
Data quality element	DQ_Element	Enumerated domain 1 – Completeness 2 – Logical consistency 3 – Positional accuracy 4 – Temporal accuracy 5 – Thematic accuracy	1 – Completeness data quality element describing the presence or absence of features, their attributes and their relationships
Data quality subelement	DQ_Subelement	Enumerated domain (Dependent upon data quality element) EXAMPLE	1 – Commission excess data in the dataset
Data quality measure	DQ_Measure		
Data quality measure description	DQ_MeasureDesc	Free text	Existence of excess items
Data quality measure identification code	DQ_MeasureID	Enumerated domain	10101
Data quality evaluation method	DQ_EvalMethod		
Data quality evaluation method type	DQ_EvalMethodType	Enumerated domain 1 – internal (direct) 2 – external (direct) 3 – indirect	2 – external
Data quality evaluation method description	DQ_EvalMethodDesc	Free text or citation (depends on data quality evaluation method type)	Compare count of items in dataset against count of items in universe of discourse
Data quality result	DQ_QualityResult		
Data quality value type	DQ_ValueType	Enumerated domain 1 – Boolean variable 2 – number 3 – ratio 4 – percentage 5 – sample 6 – table 7 – binary image 8 – matrix 9 – citation (ISO 19115) 10 – free text 11 – other	1 – Boolean variable
Data quality value	DQ_Value	Record (ISO 11404) (Depends on data quality value type)	True
Data quality value unit	DQ_ValueUnit	(Depends on data quality value)	Not applicable
Data quality date	DQ_Date	ISO 8601:1988	2000-03-05
Conformance quality level	DQ_ConformanceLevel	value or set of values	Zero difference between dataset and universe of discourse counts

[a] Short name is for use within this annex

© ISO 2003 — All rights reserved

D.3 Data quality completeness measure examples

Completeness is the presence or absence of features, their attributes and their relationships. It has the following subelements:

commission – excess data in a dataset

omission – data absent from a dataset

Table D.2 provides some examples for the subelements.

Table D.2 — Examples of data quality completeness measures

Data quality component	Example 1	Example 2	Example 3
DQ_Scope	All items classified as houses in the dataset.	All items classified as houses and bounded by longitudes –83,1 –83,3 and latitudes +38,3 +38,4.	All items classified as houses and in the town of Augusta, Georgia, U.S.A.
DQ_Element	1 – Completeness	1 – Completeness	1 – Completeness
DQ_Subelement	1 – Commission	1 – Commission	1 – Commission
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of commissions	Percentage of commissions
DQ_MeasureID	10101	10102	10103
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	Compare count of items in dataset against count of items in universe of discourse.	Compare count of items in dataset against count of items in universe of discourse.	Divide count of excess items in dataset by count of items in universe of discourse; then multiply by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	10	10
DQ_ValueUnit	NA	Houses	Percent
DQ_Date	2000-03-05	2000-03-06	2000-03-04
DQ_ConformanceLevel	Zero commissions in dataset.	Less than 9 commissions in dataset.	Less than 9 percent commissions in dataset.
Example dataset parameters	110 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.	110 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.	110 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.
Example quality result meaning	Dataset fails. Excess items exist. More items are classified as houses in the dataset than are in the universe of discourse.	Dataset fails. The number of excess items in the dataset exceeds the data quality conformance quality level.	Dataset fails. The percentage of excess items in the dataset exceeds the data quality conformance quality level.

© ISO 2003 — All rights reserved

Table D.2 *(continued)*

Data quality component	Example 4	Example 5	Example 6
DQ_Scope	All items classified as houses in the dataset.	All items classified as houses and bounded by longitudes –83,1 –83,3 and latitudes +38,3 +38,4.	All items classified as houses and in the city of Stockholm, Sweden.
DQ_Element	1 – Completeness	1 – Completeness	1 – Completeness
DQ_Subelement	2 – Omission	2 – Omission	2 – Omission
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of ommissions	Percentage of ommissions
DQ_MeasureID	10201	10202	10203
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	Compare count of items in dataset against count of items in universe of discourse.	Compare count of items in dataset against count of items in universe of discourse.	Divide count of excess items in dataset by count of items in universe of discourse; then multiply by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	10	10
DQ_ValueUnit	NA	Houses	Percent
DQ_Date	2000-03-06	2000-03-03	2000-03-04
DQ_ConformanceLevel	Zero omissions in dataset.	Less than 9 omissions in dataset.	Less than 9 percent omissions in dataset.
Example dataset parameters	90 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.	90 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.	90 items in dataset are within the data quality scope; 100 items in the universe of discourse are within the scope.
Example quality result meaning	Dataset fails. Omissions exist. Fewer items are classified as houses in the dataset than are in the universe of discourse.	Dataset fails. The number of missing items in the dataset exceeds the data quality conformance quality level.	Dataset fails. The percentage of missing items in the dataset exceeds the data quality conformance quality level.

D.4 Data quality logical consistency measure examples

Logical consistency is the degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical). Logical consistency has the following subelements:

conceptual consistency – adherence to rules of the conceptual schema

domain consistency – adherence of values to the value domain

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

Table D.3 provides some examples for the subelements.

© ISO 2003 — All rights reserved

Table D.3 — Examples of data quality logical consistency measures

Data quality component	Example 1	Example 2	Example 3
DQ_Scope	All items classified as houses in the dataset.	All items classified as houses and bounded by longitudes –83,1 –83,3 and latitudes +38,3 +38,4.	All items classified as houses and in the city of Helsinki, Finland.
DQ_Element	2 – Logical consistency	2 – Logical consistency	2 – Logical consistency
DQ_Subelement	1 – Conceptual consistency	1 – Conceptual consistency	1 – Conceptual consistency
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of violating items	Percentage of violating items
DQ_MeasureID	20101	20102	20103
DQ_EvalMethod			
DQ_EvalMethodType	1 – Internal	1 – Internal	1 – Internal
DQ_EvalMethodDesc	Count the number of features and feature relationships which violate the conceptual schema of the dataset.	Count the number of features and feature relationships which violate the conceptual schema of the dataset.	Divide count of features and feature relationships which violate the conceptual schema by count of those in dataset.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	1	1,0%
DQ_ValueUnit	NA	Count	Percent
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero violations in dataset.	Zero violations in dataset.	Zero percent violations in dataset.
Example dataset parameters	80 features and 20 feature relationships are within the data quality scope. One feature relationship exists which is not defined in the conceptual schema.	80 features and 20 feature relationships are within the data quality scope. One feature relationship exists which is not defined in the conceptual schema.	80 features and 20 feature relationships are within the data quality scope. One feature relationship exists which is not defined in the conceptual schema.
Example quality result meaning	The dataset fails. Violation of the conceptual schema exists.	The dataset fails. The number of violation items exceeds the conformance quality level.	The dataset fails. The percentage of violation items exceeds the conformance quality level.

© ISO 2003 — All rights reserved

Table D.3 *(continued)*

Data quality component	Example 4	Example 5	Example 6
DQ_Scope	All items classified as living quarters.	All items classified as living quarters and bounded by longitudes –91,3 –91,4 and latitudes +40,0 +40,2.	All items classified as living quarters and in the city of London, England, U.K.
DQ_Element	2 – Logical consistency	2 – Logical consistency	2 – Logical consistency
DQ_Subelement	2 – Domain consistency	2 – Domain consistency	2 – Domain consistency
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of domain inconsistencies	Percentage of domain inconsistencies
DQ_MeasureID	20201	20202	20203
DQ_EvalMethod			
DQ_EvalMethodType	1 – Internal	1 – Internal	1 – Internal
DQ_EvalMethodDesc	Compare attribute of items within scope against acceptable attribute domain (acceptable attributes) and determine if any are outside the domain.	Compare attribute of items within scope against acceptable attribute domain (acceptable attributes) and count those that are outside the domain.	Divide the number of items with attribute violations by the total number of items within scope and multiply the result by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	8	8,0%
DQ_ValueUnit	NA	Attribute violations	Percent
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items with attribute violations.	10 or fewer items with attribute violations.	Less than 5,0 percent of the items with attribute violations.
Example dataset parameters	100 items within scope in the dataset. Eight of the items have attributes that violate the attribute domain.	100 items within scope in the dataset. Eight of the items have attributes that violate the attribute domain.	100 items within scope in the dataset. Eight of the items have attributes that violate the attribute domain.
Example quality result meaning	Dataset fails. The attributes of at least one item violated the attribute domain.	Dataset passes. Fewer than ten items had attributes that violated the attribute domain.	Dataset fails. Greater than 5 percent of the items had attributes that violated the attribute domain.

© ISO 2003 — All rights reserved

Table D.3 *(continued)*

Data quality component	Example 7	Example 8	Example 9
DQ_Scope	All records in the dataset for items classified as living quarters.	All records in the dataset for items classified as living quarters and bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All records in the dataset for items classified as living quarters and in the city of Tokyo, Japan.
DQ_Element	2 – Logical consistency	2 – Logical consistency	2 – Logical consistency
DQ_Subelement	3 – Format consistency	3 – Format consistency	3 – Format consistency
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of format inconsistencies	Percentage of format inconsistencies
DQ_MeasureID	20301	20302	20303
DQ_EvalMethod			
DQ_EvalMethodType	1 – Internal	1 – Internal	1 – Internal
DQ_EvalMethodDesc	Compare the record structure for all items within scope to specified field definitions and structure and count those that are inconsistent. Specifically assure that the field for the living quarter's type code is an alphabetic field of 5 characters in length.	Compare the record structure for all items within scope to specified field definitions and structure and count those that are inconsistent. Specifically assure that the field for the living quarter's type code is an alphabetic field of 5 characters in length.	Compare the record structure for all items within scope to specified field definitions and structure and count those that are inconsistent. Specifically assure that the field for the living quarter's type code is an alphabetic field of 5 characters in length. Divide the count by the number of records checked and multiply the results by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	8	8,0%
DQ_ValueUnit	NA	Format violations	Percent
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items may have format violations.	Zero items may have format violations.	Zero percent of the items may have format violations.
Example dataset parameters	100 items within scope in the dataset. Eight of the items violate the specified format.	100 items within scope in the dataset. Eight of the items violate the specified format.	100 items within scope in the dataset. Eight of the items violate the specified format.
Example quality result meaning	Dataset fails. Format violations found.	Dataset fails. Format violations found.	Dataset fails. Format violations found.

© ISO 2003 — All rights reserved

Table D.3 *(continued)*

Data quality component	Example 10	Example 11	Example 12
DQ_Scope	All province boundaries in the dataset.	All state boundaries in the United States.	All state boundaries in the United States.
DQ_Element	2 – Logical consistency	2 – Logical consistency	2 – Logical consistency
DQ_Subelement	4 – Topological consistency	4 – Topological consistency	4 – Topological consistency
DQ_Measure			
DQ_MeasureDesc	Pass-Fail	Number of items with topological inconsistencies	Percentage of items with topological inconsistencies
DQ_MeasureID	20401	20402	20403
DQ_EvalMethod			
DQ_EvalMethodType	1 – Internal	1 – Internal	1 – Internal
DQ_EvalMethodDesc	For each province, check the boundaries to assure closure. Count the number of provinces whose boundaries do not close.	For each state, check the boundaries to assure closure. Count the number of provinces whose boundaries do not close.	For each state, check the boundaries to assure closure. Count the number of provinces whose boundaries do not close. Divide the count by the number of records checked and multiply the result by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	2	2,0%
DQ_ValueUnit	NA	Topological inconsistencies	Percent of topological inconsistencies
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items may have topological violations.	Zero items may have topological violations.	Zero percent of the items may have topological violations.
Example dataset parameters	100 items within scope in the dataset. Two of the items have topological inconsistencies.	100 items within scope in the dataset. Two of the items have topological inconsistencies.	100 items within scope in the dataset. Two of the items have topological inconsistencies.
Example quality result meaning	Dataset fails. Topological inconsistencies found.	Dataset fails. Number of topological inconsistencies exceeds conformance quality level.	Dataset fails. Percentage of topological inconsistencies exceeds conformance quality level.

D.5 Data quality positional accuracy measure examples

Positional accuracy is the accuracy of the position of a feature. Positional accuracy has the following subelements:

absolute or external accuracy – closeness of reported coordinate values to values accepted as or being true

relative or internal accuracy – closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true

gridded data position accuracy – closeness of gridded data position values to values accepted as or being true

© ISO 2003 — All rights reserved

Table D.4 provides some examples for the subelements.

Table D.4 — Examples of data quality positional accuracy measures

Data quality component	Example 1	Example 2
DQ_Scope	All nodes forming road boundaries in the dataset.	All nodes forming road boundaries in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.
DQ_Element	3 – Positional accuracy	3 – Positional accuracy
DQ_Subelement	1 – Absolute or external accuracy	1 – Absolute or external accuracy
DQ_Measure		
DQ_MeasureDesc	RMSE	Percentage of items with coordinate error greater than specification limit.
DQ_MeasureID	30101	30102
DQ_EvalMethod		
DQ_EvalMethodType	2 – External	2 – External
DQ_EvalMethodDesc	For each node, measure the error distance between absolute coordinate values of the node in the dataset and those in the universe of discourse. Compute RMSE from the error distances.	For each node, measure the error distance between absolute coordinate values of the node in the dataset and those in the universe of discourse. Count the number of the nodes whose error distance exceeds the specification limit (e.g.1 m). Divide the number of the non-conforming nodes by the number of the nodes in the data quality scope. Multiply the result by 100.
DQ_QualityResult		
DQ_ValueType	2 – Number	4 – Percentage
DQ_Value	1,70 m	25%
DQ_ValueUnit	Metre	Percentage
DQ_Date	2000-03-06	2000-03-06
DQ_ConformanceLevel	Not specified	Not specified
Example dataset parameters	Omitted.	Omitted.
Example quality result meaning	RMSE of distance of the nodes is 1,70 m. Since conformance quality level is not specified, only the RMSE is reported.	25% of the nodes within the data quality scope have error distance more than 1 m. Since conformance quality level is not specified, only the percentage is reported.

© ISO 2003 — All rights reserved

Table D.4 *(continued)*

Data quality component	Example 3	Example 4
DQ_Scope		
DQ_Element	3 – Positional accuracy	3 – Positional accuracy
DQ_Subelement	2 – Relative or internal accuracy	2 – Relative or internal accuracy
DQ_Measure		
DQ_MeasureDesc	RMSE	Percentage of items with coordinate error greater than specification limit.
DQ_MeasureID	30201	30202
DQ_EvalMethod		
DQ_EvalMethodType	2 – External	2 – External
DQ_EvalMethodDesc	For each node, measure the error distance between relative coordinate values of the node in the dataset and those in the universe of discourse. Compute RMSE from the error distances.	For each node, measure the error distance between relative coordinate values of the node in the dataset and those in the universe of discourse. Count the number of the nodes whose error distance exceeds the specification limit (e.g.1 m). Divide the number of the non-conforming nodes by the number of the nodes in the data quality scope. Multiply the result by 100.
DQ_QualityResult		
DQ_ValueType	2 – Number	4 – Percentage
DQ_Value	1,50 m	20%
DQ_ValueUnit	Metre	Percentage
DQ_Date	2000-03-06	2000-03-06
DQ_ConformanceLevel	Not specified	Not specified
Example dataset parameters	Omitted.	Omitted.
Example quality result meaning	RMSE of distance of the nodes is 1,50 m. Since conformance quality level is not specified, only the RMSE is reported.	20% of the nodes within the data quality scope have error distance more than 1 m. Since conformance quality level is not specified, only the percentage is reported.

 © ISO 2003 — All rights reserved

Table D.4 *(continued)*

Data quality component	Example 5	Example 6	Example 7
DQ_Scope	All gridded elevation point data of DEM in the dataset.	All gridded elevation point data of DEM in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All gridded elevation point data of DEM in the city of Bangkok, Thailand.
DQ_Element	3 – Positional accuracy	3 – Positional accuracy	3 – Positional accuracy
DQ_Subelement	3 – Gridded data position accuracy	3 – Gridded data position accuracy	3 – Gridded data position accuracy
DQ_Measure			
DQ_MeasureDesc	RMSE	Percentage of items with coordinate error greater than specification limit.	Pass-fail
DQ_MeasureID	30301	30302	30303
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	For each gridded point, measure the difference between absolute height value of the point in the dataset and that in the universe of discourse. Compute RMSE from the height differences.	For each gridded point, measure the difference between absolute height value of the point in the dataset and that in the universe of discourse. Count the number of the points whose height difference exceeds the specification limit (e.g.1 m). Divide the number of the non-conforming points by the number of the points in the data quality scope. Multiply the result by 100.	For each gridded point, measure the difference between absolute height value of the point in the dataset and that in the universe of discourse. Count the number of the points whose height difference exceeds the specification limit (e.g.1 m). Divide the number of the non-conforming points by the number of the points in the data quality scope and multiply the ratio by 100. Compare the percentage of the non-conforming points against the conformance quality level.
DQ_QualityResult			
DQ_ValueType	2 – Number	4 – Percentage	1 – Boolean variable
DQ_Value	0,8 m	8%	False
DQ_ValueUnit	Metre	Percent of points with height error greater than the specification limit.	NA
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Not specified	Not specified	Less than 5 percent of items may have height error greater than specification limit.
Example dataset parameters	Omitted	Omitted	Omitted
Example quality result meaning	RMSE of height is 0,8 m. Since conformance quality level is not specified, only the RMSE is reported.	8 percent of the gridded points within the data quality scope have height error more than 1 m. Since conformance quality level is not specified, only the percentage is reported.	Dataset fails. Percentage of non conforming points exceeds the conformance quality level.

© ISO 2003 — All rights reserved

D.6 Data quality temporal accuracy measure examples

Temporal accuracy is the accuracy of the temporal attributes and temporal relationships. Temporal accuracy has the following subelements:

accuracy of a time measurement – correctness of the temporal references of an time (reporting of error in time measurement)

temporal consistency – correctness of ordered events or sequences

temporal validity – validity of data with respect to time

Table D.5 provides some examples of thematic accuracy for the subelements accuracy of a time measurement, temporal consistency and temporal validity.

© ISO 2003 — All rights reserved

Table D.5 — Examples of data quality temporal accuracy measures

Data quality component	Example 1	Example 2	Example 3
DQ_Scope	All traffic accident data in the dataset.	All traffic accident data in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All traffic accident data in the city of London, UK.
DQ_Element	4 – Temporal accuracy	4 – Temporal accuracy	4 – Temporal accuracy
DQ_Subelement	1 – Accuracy of a time measurement	1 – Accuracy of a time measurement	1 – Accuracy of a time measurement
DQ_Measure			
DQ_MeasureDesc	RMSE	Percentage of items with the error of temporal attribute greater than specification limit.	Pass-fail
DQ_MeasureID	40101	40102	40103
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	For each traffic accident data, measure the difference between accident occurrence time in the dataset and that in the universe of discourse. Compute RMSE from the occurrence time differences.	For each traffic accident data, measure the difference between accident occurrence time in the dataset and that in the universe of discourse. Count the number of the accidents whose occurrence time difference exceeds the specification limit (e.g. 2 h). Divide the number of the non-conforming accident data by the number of accident data in the data quality scope, and multiply the result by 100.	For each traffic accident data, measure the difference between accident occurrence time in the dataset and that in the universe of discourse. Count the number of the accidents whose occurrence time difference exceeds the specification limit (e.g. 2 h). Divide the number of the non-conforming accident data by the number of accident data in the data quality scope, and multiply the result by 100. Compare the percentage of the non-conforming accident data against the conformance quality level.
DQ_QualityResult			
DQ_ValueType	2 – Number	4 – Percentage	1 – Boolean variable
DQ_Value	1,5 h	18	False
DQ_ValueUnit	Hours	Percent	NA
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Not specified	Not specified	10%
Example dataset parameters	Omitted	Omitted	Omitted
Example quality result meaning	RMSE of occurrence time is 1,5 h. Since conformance quality level is not specified, only the RMSE is reported.	18% of the accident data within the data quality scope have occurrence time error more than 2 h. Since conformance quality level is not specified, only the percentage is reported.	Dataset fails. Percentage of non conforming accident data exceeds the conformance quality level.

© ISO 2003 — All rights reserved

Table D.5 *(continued)*

Data quality component	Example 4	Example 5	Example 6
DQ_Scope	All historical event data in the dataset.	All historical event data in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All historical event data in China.
DQ_Element	4 – Temporal accuracy	4 – Temporal accuracy	4 – Temporal accuracy
DQ_Subelement	2 – Temporal consistency	2 – Temporal consistency	2 – Temporal consistency
DQ_Measure			
DQ_MeasureDesc	Pass-fail	Number of items with inconsistent temporal relationships	Percent of items with inconsistent temporal relationships
DQ_MeasureID	40201	40202	40203
DQ_EvalMethod			
DQ_EvalMethodType	1 – External	1 – External	1 – External
DQ_EvalMethodDesc	Check each historical event to assure that it is correctly ordered against the rest of event data.	Check each historical event to assure that it is correctly ordered against the rest of event data. Count those that are not correctly ordered.	Check each historical event to assure that it is correctly ordered against the rest of event data. Count those that are not correctly ordered. Divide the result by the total number of items within scope and multiply it by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	3	60%
DQ_ValueUnit	NA	Temporal inconsistencies	Percent of temporal inconsistencies
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items may have temporal inconsistency.	Zero items may have temporal inconsistency.	Zero percent of items may have temporal inconsistency.
Example dataset parameters	5 historical events in the data quality scope; {A,B,C,D,E} is the correct sequence. In the dataset, the five events are recorded in the order of {A,B,D,E,C}. Individual event (A,B,C,D,E) is defined to be an item. Items with inconsistent order are (C,D,E).	5 historical events in the data quality scope; {A,B,C,D,E} is the correct sequence. In the dataset, the five events are recorded in the order of {A,B,D,E,C}. Individual event (A,B,C,D,E) is defined to be an item. Items with inconsistent temporal order are (C,D,E).	5 historical events in the data quality scope; {A,B,C,D,E} is the correct sequence. In the dataset, the five events are recorded in the order of {A,B,D,E,C}. Individual event (A,B,C,D,E) is defined to be an item. Items with inconsistent order are (C,D,E).
Example quality result meaning	Dataset fails. Temporal inconsistency is found.	Dataset fails. Number of temporal inconsistencies exceeds conformance quality level.	Dataset fails. Percentage of temporal inconsistencies exceeds conformance quality level.

© ISO 2003 — All rights reserved

Table D.5 *(continued)*

Data quality component	Example 7	Example 8	Example 9
DQ_Scope	All land price data in the dataset.	All land price data in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All land price data in the city of Tokyo, Japan.
DQ_Element	4 – Temporal accuracy	4 – Temporal accuracy	4 – Temporal accuracy
DQ_Subelement	3 – Temporal validity	3 – Temporal validity	3 – Temporal validity
DQ_Measure			
DQ_MeasureDesc	Pass-fail	Number of items with temporal invalidity	Percentage of items with temporal invalidity
DQ_MeasureID	40301	40302	40303
DQ_EvalMethod			
DQ_EvalMethodType	1 – Internal	1 – Internal	1 – Internal
DQ_EvalMethodDesc	Check land price data to assure that it was surveyed in 1995.	Check land price data to assure that it was surveyed in 1995. Count those that were not surveyed in 1995.	Check land price data to assure that it was surveyed in 1995. Count those that were not surveyed in 1995. Divide the result by the total number of items in data quality scope and multiply it by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	5	5%
DQ_ValueUnit	NA	Temporal invalidity	Percent
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items may have temporal invalidity.	10 or fewer items may have temporal invalidity.	Less than 10 percent of the items may have temporal invalidity.
Example dataset parameters	100 items with the collection date of 1995 in the dataset; 95 were actually collected in 1995; 5 were actually collected in 1985.	100 items with the collection date of 1995 in the data quality scope; 95 were actually collected in 1995; 5 were actually collected in 1985.	100 items with the collection date of 1995 in the data quality scope; 95 were actually collected in 1995; 5 were actually collected in 1985.
Example quality result meaning	Dataset fails. At least one item has temporal invalidity.	Dataset passes. Fewer than ten items had temporal invalidity.	Dataset passes. Less than 10 percent of the items had temporal invalidity.

D.7 Data quality thematic accuracy measure examples

Thematic accuracy is the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. Thematic accuracy has the following subelements:

classification correctness – comparison of the classes assigned to features or their attributes to a universe of discourse (ground truth or reference dataset)

non-quantitative attribute correctness – correctness of non-quantitative attributes

quantitative attribute correctness – accuracy of quantitative attributes

© ISO 2003 — All rights reserved

Table D.6 provides some examples of thematic accuracy for the subelements.

Table D.6 — Examples of data quality thematic accuracy measures

<table>
<tr><th>Data quality component</th><th>Example 1</th><th>Example 2</th><th>Example 3</th></tr>
<tr><td>DQ_Scope</td><td>All items classified as A, B and C in the dataset.</td><td>All items classified as A, B and C in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.</td><td>All items classified as A, B and C in Saudi Arabia.</td></tr>
<tr><td>DQ_Element</td><td>5 – Thematic accuracy</td><td>5 – Thematic accuracy</td><td>5 – Thematic accuracy</td></tr>
<tr><td>DQ_Subelement</td><td>1 – Classification correctness</td><td>1 – Classification correctness</td><td>1 – Classification correctness</td></tr>
<tr><td>DQ_Measure</td><td></td><td></td><td></td></tr>
<tr><td>DQ_MeasureDesc</td><td>Pass-fail</td><td>Percent correctly classified (PCC)</td><td>Misclassification percentage matrix</td></tr>
<tr><td>DQ_MeasureID</td><td>50101</td><td>50102</td><td>50103</td></tr>
<tr><td>DQ_EvalMethod</td><td></td><td></td><td></td></tr>
<tr><td>DQ_EvalMethodType</td><td>2 – External</td><td>2 – External</td><td>2 – External</td></tr>
<tr><td>DQ_EvalMethodDesc</td><td>For each item in the dataset, compare the assigned class against true class in the universe of discourse.</td><td>For each item in the data quality scope, compare the assigned class against true class in the universe of discourse. Count items which are correctly classified. Divide the results by the total number of the items in data quality scope and multiply it by 100.</td><td>For each item in the data quality scope, compare the assigned class against true class in the universe of discourse. Generate a matrix of N(i,j) where N(i,j) is the number of items of class (i) which are classified as class (j) in the dataset. Divide N(i,j) by the total number of items of class (i) and multiply them by 100.</td></tr>
<tr><td>DQ_QualityResult</td><td></td><td></td><td></td></tr>
<tr><td>DQ_ValueType</td><td>1 – Boolean variable</td><td>4 – Percentage</td><td>8 – Matrix</td></tr>
<tr><td>DQ_Value</td><td>False</td><td>60%</td><td><table>
<tr><td></td><td colspan="5">Dataset class</td></tr>
<tr><td rowspan="5">True class i</td><td></td><td>A</td><td>B</td><td>C</td><td>%</td></tr>
<tr><td>A</td><td>70</td><td>20</td><td>10</td><td>100</td></tr>
<tr><td>B</td><td>20</td><td>40</td><td>40</td><td>100</td></tr>
<tr><td>C</td><td>20</td><td>20</td><td>60</td><td>100</td></tr>
<tr><td>%</td><td>100</td><td>100</td><td>100</td><td></td></tr>
</table></td></tr>
<tr><td>DQ_ValueUnit</td><td>NA</td><td>Percent</td><td>Percent</td></tr>
<tr><td>DQ_Date</td><td>2000-03-06</td><td>2000-03-06</td><td>2000-03-06</td></tr>
<tr><td>DQ_ConformanceLevel</td><td>Zero items may have classification errors.</td><td>Greater than 80 percent of the items shall be correctly classified.</td><td>Not specified.</td></tr>
<tr><td></td><td><table>
<tr><td></td><td colspan="5">Dataset class</td></tr>
<tr><td rowspan="5">True class</td><td></td><td>A</td><td>B</td><td>C</td><td>Count</td></tr>
<tr><td>A</td><td>7</td><td>2</td><td>1</td><td>10</td></tr>
<tr><td>B</td><td>1</td><td>2</td><td>2</td><td>5</td></tr>
<tr><td>C</td><td>1</td><td>1</td><td>3</td><td>5</td></tr>
<tr><td>Count</td><td>9</td><td>5</td><td>6</td><td>20</td></tr>
</table></td><td><table>
<tr><td></td><td colspan="5">Dataset class</td></tr>
<tr><td rowspan="5">True class</td><td></td><td>A</td><td>B</td><td>C</td><td>Count</td></tr>
<tr><td>A</td><td>7</td><td>2</td><td>1</td><td>10</td></tr>
<tr><td>B</td><td>1</td><td>2</td><td>2</td><td>5</td></tr>
<tr><td>C</td><td>1</td><td>1</td><td>3</td><td>5</td></tr>
<tr><td>Count</td><td>9</td><td>5</td><td>6</td><td>20</td></tr>
</table></td><td><table>
<tr><td></td><td colspan="5">Dataset class</td></tr>
<tr><td rowspan="5">True class</td><td></td><td>A</td><td>B</td><td>C</td><td>Count</td></tr>
<tr><td>A</td><td>7</td><td>2</td><td>1</td><td>10</td></tr>
<tr><td>B</td><td>1</td><td>2</td><td>2</td><td>5</td></tr>
<tr><td>C</td><td>1</td><td>1</td><td>3</td><td>5</td></tr>
<tr><td>Count</td><td>9</td><td>5</td><td>6</td><td>20</td></tr>
</table></td></tr>
<tr><td>Example quality result meaning</td><td>Dataset fails. 8 items are misclassified.</td><td>Dataset fails. 40 percent of the items are misclassified.</td><td>Since no conformance quality level is specified, misclassification matrix is reported.</td></tr>
</table>

© ISO 2003 — All rights reserved

Table D.6 *(continued)*

Data quality component	Example 4	Example 5	Example 6
DQ_Scope	All items with geographic names in the dataset.	All items with geographic names in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All items with geographic names in the city of Lisbon, Portugal.
DQ_Element	5 – Thematic accuracy	5 – Thematic accuracy	5 – Thematic accuracy
DQ_Subelement	2 – Non-quantitative attribute correctness	2 – Non-quantitative attribute correctness	2 – Non-quantitative attribute correctness
DQ_Measure			
DQ_MeasureDesc	Pass-fail	Number of items with incorrect geographic names	Percentage of items with incorrect geographic names
DQ_MeasureID	50201	50202	50203
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	Compare the geographic names in the dataset against those in the universe of discourse.	Compare the geographic names in the data quality scope against those in the universe of discourse. Count items with incorrect geographic names.	Compare the geographic names in the data quality scope against those in the universe of discourse. Count items with incorrect geographic names. Divide the result by the total number of items in the data quality scope and multiply it by 100.
DQ_QualityResult			
DQ_ValueType	1 – Boolean variable	2 – Number	4 – Percentage
DQ_Value	False	5	5%
DQ_ValueUnit	NA	Number of items with incorrect geographic names	Percent
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Zero items may have incorrect geographic names.	Less than 3 items may have incorrect geographic names.	Less than 3 percent of items may have incorrect geographic names.
Example dataset parameters	100 items with geographic name in the dataset; 5 names are misspelled.	100 items with geographic name in the data quality scope; 5 names are misspelled.	100 items with geographic name in the data quality scope; 5 names are misspelled.
Example quality result meaning	Dataset fails. At least one item has incorrect geographic names.	Dataset fails. Greater than 3 items have incorrect geographic names.	Dataset fails. Greater than 3 percent of the items have incorrect geographic names.

© ISO 2003 — All rights reserved

Table D.6 *(continued)*

Data quality component	Example 7	Example 8	Example 9
DQ_Scope	All items which have temperature attribute in the dataset.	All items which have temperature attribute in the area bounded by longitudes +139 +140 and latitudes +36,0 +37,0.	All items which have temperature attribute in the city of Munich, Germany.
DQ_Element	5 – Thematic accuracy	5 – Thematic accuracy	5 – Thematic accuracy
DQ_Subelement	3 – Quantitative attribute accuracy	3 – Quantitative attribute accuracy	3 – Quantitative attribute accuracy
DQ_Measure			
DQ_MeasureDesc	RMSE	Percentage of items with temperature error greater than specification limit	Pass-fail
DQ_MeasureID	50301	50302	50303
DQ_EvalMethod			
DQ_EvalMethodType	2 – External	2 – External	2 – External
DQ_EvalMethodDesc	For each item, measure the difference between temperature value in the dataset and that in the universe of discourse. Compute RMSE from the differences.	For each item, measure the difference between temperature value in the dataset and that in the universe of discourse. Count the number of the items whose temperature difference exceeds the specification limit (e.g.1 deg.). Divide the number of the non-conforming items by the number of the items in the data quality scope. Multiply the result by 100.	For each item, measure the difference between temperature value in the dataset and that in the universe of discourse. Count the number of the items whose temperature difference exceeds the specification limit (e.g.1 deg.). Divide the number of the non-conforming items by the number of the items in the data quality scope. Multiply the result by 100. Compare the percentage of the non-conforming items against the conformance quality level.
DQ_QualityResult			
DQ_ValueType	2 – Number	4 – Percentage	1 – Boolean variable
DQ_Value	0,5	5	False
DQ_ValueUnit	Degree	Percent of the items with temperature error greater than the specification limit	NA
DQ_Date	2000-03-06	2000-03-06	2000-03-06
DQ_ConformanceLevel	Not specified	Not specified	Less than 1 percent of items may have temperature error greater than specification limit.
Example dataset parameters	Omitted	Omitted	Omitted
Example quality result meaning	RMSE of temperature is 0,5 deg. Since conformance quality level is not specified, only the RMSE is reported.	5 percent of the items within the data quality scope have temperature error more than 1 deg. Since conformance quality level is not specified, only the percentage is reported.	Dataset fails. Percentage of non conforming items exceeds the conformance quality level.

© ISO 2003 — All rights reserved

Annex E
(informative)

Guidelines for sampling methods applied to geographic datasets

E.1 Introduction

This annex provides guidelines for defining samples and devising sampling methods. For sampling for evaluating conformance to a product specification, ISO 2859 and ISO 3951 may be applied. These standards were originally developed for non-spatial use. This annex describes how to apply ISO 2859 and ISO 3951 and other spatial sampling techniques to geographic information.

E.2 Lot and item

The lot and item are important concepts in the sampling inspection method specified by ISO 2859 and ISO 3951. A lot is the minimum unit for which quality may be evaluated. An item is the minimum unit to be inspected and should be defined by the data producer in accordance with the product specification.

E.3 Sample size

The size of a population, and consequently the size of samples, may be defined according to different bases on items. The definition of a sample size requires an explicit indication of the items. Examples of different bases are presented in Table E.1.

The difference between the perspectives is illustrated in Figure E.1. The whole figure represents the data within the data quality scope. The figure depicts a possible sample area of approximately 15% of the total data quality scope area, but only about 10% of the curve length within the sample area, and 0% of the vertices.

To help overcome sample difficulties such as those in Figure E.1, the size and location of a sample might be defined using a combination of different criteria, thus enforcing the representativeness of the sample.

EXAMPLE The sample should include 10% of the area covered by the dataset and contain not less than 5% of the total curve length describing the objects in the dataset.

Table E.1 — Different bases for defining population

Basis	Size of the dataset	Sample size
Features	Number of features of a given type	Number of features of a given type expressed as percentage of the total number of objects
Area covered	Area covered by the dataset	Area covered by the sample expressed as percentage of the total area
Curves	Total length of the curves in the dataset	Length of the sampled curves expressed as a percentage of the total length
Vertices	Total number of vertices describing curves or areas in the dataset	Number of vertices in the sample expressed as a percentage of the total number of vertices

© ISO 2003 — All rights reserved

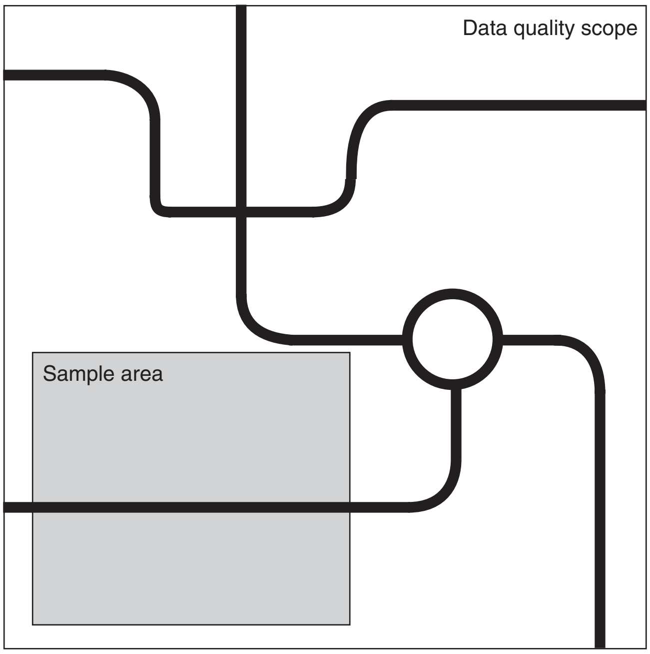


Figure E.1 — Effect of sample area location on representativeness of items in the sample

E.4 Sampling strategies

E.4.1 Introduction

This section provides guidelines for defining samples and sampling methods, considering particular aspects of geographic data. The sampling strategies described in this annex are shown graphically in Figure E.2. There are two aspects to a sampling strategy, the items to be sampled (area or feature) and the manner by which the items are selected (probability or judgement).

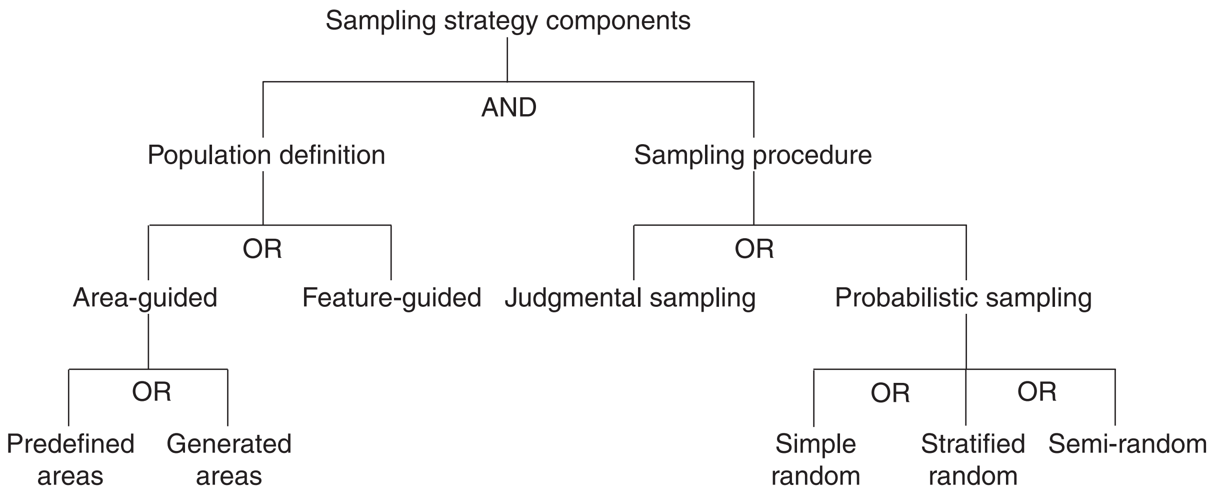


Figure E.2 — Sampling strategy relationships

© ISO 2003 — All rights reserved

E.4.2 Probabilistic versus judgemental sampling

E.4.2.1 Differences

Probabilistic sampling applies sampling theory and involves random selection of the sample items. The essential characteristic of probabilistic sampling is that each member of the population from which the sample is selected has a known probability of selection. When probabilistic sampling is used, statistical inferences may be made about the sampled population. Judgemental sample designs involve selection of samples based on expert knowledge or professional judgement.

E.4.2.2 Simple random sampling

Simple random sampling is probability-based and involves selection of samples randomly. The particular sample (e.g. features, location, time) is selected using random numbers to identify the items and all possible selections are equally likely. Simple random sampling is useful when the population of interest is relatively homogeneous in the characteristics being sampled, i.e. no major patterns and clusters. This method may not result in representative coverage of an area, i.e., it is possible that the sample selected will be only from a part of the area.

E.4.2.3 Stratified random sampling

Stratified sampling requires the population to be separated into non-overlapping strata or subpopulations that are more homogeneous among sample items in the same strata than among sample items in different strata. This sampling strategy has the potential for greater precision in estimates of mean and variance than that of a non-stratified strategy for the same population.

E.4.2.4 Semi-random sampling

Semi-random or systematic sampling applies random selection of the initial sample items (e.g. location, time, feature) and rules for selection for all remaining items. An example of semi-random or systematic sampling is grid sampling where the initial position of a grid is randomly determined and samples are taken at regularly spaced intervals (grid cells) over space. Systematic grid sampling is used to search for clusters and to infer means, percentiles, or other parameters and is useful for estimating spatial trends or patterns. This method provides a practical and easy way to ensure coverage of an area.

E.4.3 Feature-guided versus area-guided sampling

E.4.3.1 Feature-guided sampling (non spatial sampling)

A feature-guided sampling strategy selects sample items based on the non-spatial attributes of the features and not on their spatial location. A sample within a data quality scope can be selected randomly, assuming homogeneous production characteristics for the entire data quality scope. In some cases, simple random sampling may not produce a satisfactory sample, because homogeneity may be found only for subsets and homogeneous distribution of samples may be required, i.e., major patterns or clusters occur in the characteristics being sampled. In that case, a stratified or semi-random sampling may give better results.

EXAMPLE 1 If the sampling method is defined by selecting features randomly, then there is the risk of the occurrence of a sample being concentrated in a small area (which may not be acceptable).

Semi-random sampling may be used to ensure the verification of different criteria on the sample size and/or location, to satisfy supplementary constraints for the samples or to reduce costs of the inspection process.

EXAMPLE 2 A power company needs to evaluate the correctness of the attributes surveyed for features of different types. Two methods were considered: a random selection and a semi-random selection (selecting randomly the features of one type and then collecting the objects of different types in the neighbourhood of the first one until the samples for each type become fulfilled) leading to a reduced field inspection cost.

© ISO 2003 — All rights reserved

E.4.3.2 Area-guided sampling (spatial sampling)

In an area-guided sampling strategy, selection of sampling units is based on spatial considerations. The sampling units may be existing geographic areas (e.g. political or statistical areas) or some other partitioning of the universe of discourse for which the inspection is conducted. This type of sampling may be used as a first stage of sampling, followed by a feature-guided sampling within each subarea.

EXAMPLE 1 Random selection of UTM 1x1 km grid areas in order to evaluate the attributes of the objects contained in that area.

EXAMPLE 2 Figure E.3 illustrates the result of the definition of areas to be submitted for inspection, obtained by random generation of centre point coordinates of squares of equal area (constrained to be non-overlapping).

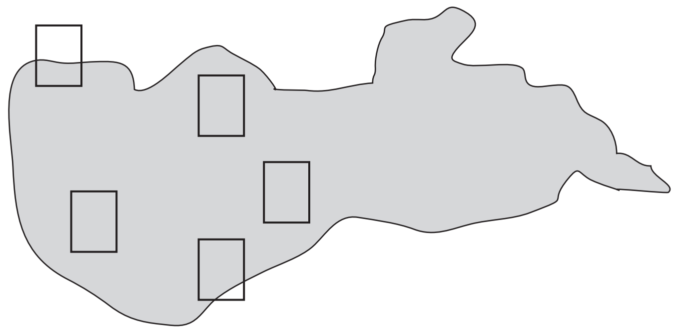


Figure E.3 — Example of area-guided random sampling

When coverage of the entire area is important, then the sample locations should be determined according to a regular or semi-regular pattern. Figure E.4 illustrates an example of semi-random (systematic) sampling with the sampled features distributed along a regular pattern used to evaluate the positional accuracy of a dataset.

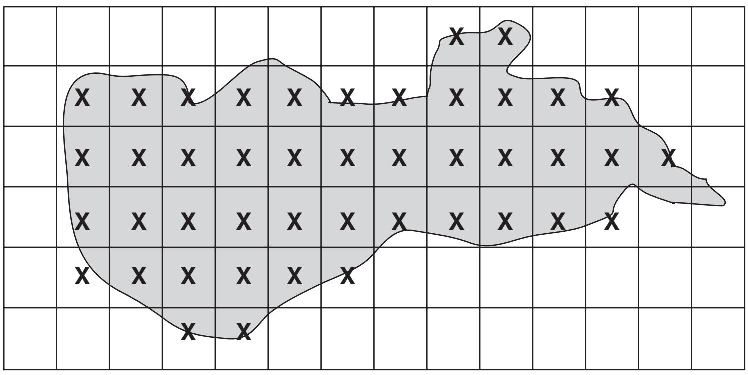


Figure E.4 — Example of area-guided regular and non-random sampling (The X's denote the grid cells selected by rule for inclusion in the sample)

Spatial partitioning with different sizes in different areas of the dataset may be needed in semi-random sampling, if the distribution of features is non-homogeneous. When using a grid of constant cell size, a rule is needed to include or exclude cells that are not completely inside the area of interest.

© ISO 2003 — All rights reserved

E.5 Probability-based sampling

E.5.1 General considerations

In applying sampling, the following points need to be taken into account:

a) The areas covered by a geographic dataset may form a continuous space. When splitting the dataset into lots, special attention should be paid to the omission or commission of items crossing over the lot boundaries.

b) A variety of factors, including the quality of source data and skill of operators, may affect the quality of geographic data. The data producer should be careful to define lots to achieve homogeneity in terms of quality.

E.5.2 Existing standard for inspection by sampling

Based on the characteristics of production and in accordance with the product specification, suitable ISO standards for inspection by sampling should be selected from the existing ISO standards. ISO 2859-1 is primarily for the inspection of a continuing series of lots. ISO 2859-2 may be applied for individual or isolated lots, while ISO 2859-3 is applied for skip-lot sampling procedures. ISO 3951 is for the inspection by variables for percentage non-conforming items.

The conformance quality level of a dataset is specified as AQL (acceptable quality level) in the case of ISO 2859-1, ISO 2859-3 and ISO 3951, and LQ (limiting quality) in the case of ISO 2859-2 based on the product specification.

Specification limits for determining conformity of each item should be specified when applying the ISO 2859 series based on the product specification. In applying ISO 3951, quality statistics (Q) should be specified based on the product specification.

E.5.3 The process of sampling

E.5.3.1 Define items

Items should be defined according to the product specification or requirements. If non-conforming items are statistically highly correlated, they are handled as a single item.

E.5.3.2 Define data quality scope of a dataset to be inspected

If the data quality scope is not homogeneous, it should be divided into homogeneous subsets. These homogeneous subsets should be considered as separate data quality scopes.

NOTE Homogeneity may be deduced where the following conditions occur:

- source data of production have almost the same quality;
- production systems (hardware, software, skill of operator) are essentially the same;
- other factors which may affect the likelihood of non-conformance occurrence such as complexity and density of features are essentially the same.

E.5.3.3 Divide the data quality scope into lots

Lots are generated by dividing the data quality scope. When there is a strong positive spatial auto-correlation of the occurrence of non-conformance, smaller lot size is desirable.

© ISO 2003 — All rights reserved

E.5.3.4 Divide the lot into sampling unit

A sampling unit may be an existing geographic area or some other partitioning of the universe of discourse for which the inspection is conducted. When the sampling unit is a geographic area, rules must be provided for inclusion of items partially in a sampling unit must be provided.

E.5.3.5 Select sampling units by simple random sampling for inspection

The total number of items which belong to selected sampling units should be as specified by relevant ISO standards.

NOTE If lots are statistically heterogeneous, simple random sampling with the same level of sampling may not be applied. ISO 2859 additionally allows for stratified sampling.

E.5.3.6 Inspection of selected sampling units

All items which belong to the selected sampling units are inspected. The items in the dataset are compared with the universe of discourse according to the chosen quality measure.

© ISO 2003 — All rights reserved

Annex F
(informative)

Example of testing for thematic accuracy and completeness

F.1 Introduction

This example is based upon the techniques used by a national land survey in Europe. The national land survey is producing a Topographic Database (TDB). This is used to make printed topographic maps at a scale of 1:10 000. The TDB also is used to make several generalised datasets. The quality conformance levels have been defined in the product specification.

The objective of this example is to illustrate a quality evaluation procedure in use for measurement of thematic accuracy and completeness in a national topographic dataset. Positional accuracy is not discussed because it is not the subject of this type of report, however, in general, positional accuracy is also tested by field surveys using non-random sampling.

The data producer's quality evaluation process is explained in F.2 and the reporting of quality information in F.6 according to this International Standard.

F.2 Quality evaluation process

Table F.1 shows the operations of the quality evaluation process of this example.

Table F.1 — Quality evaluation process

Process step	Example 1	Example 2
Identify an applicable data quality element	Completeness	Thematic accuracy
Identify an applicable data quality subelement	Commission and omission	Classification correctness
Identify a data quality scope	Topographic database /selected datasets (1:10 000 map sheets)	
Identify a data quality measure	Conformance/number of errors	Conformance/number of errors
Select and apply a data quality evaluation method	External direct quality evaluation	External direct quality evaluation
Describe sampling method	Multistage sampling	Multistage sampling
Specify conformance quality level	AQL=4	AQL=4
Determine quantitative data quality result	see F.4 and Table F.3	
Assess conformance to product specification	see F.5 and Table F.4	
Report quality evaluation results	see F.6	

© ISO 2003 — All rights reserved

F.3 Method for data quality evaluation

F.3.1 Sampling procedure

Completeness and thematic accuracy testing is carried out by applying the principles of ISO 2859-1, *Sampling procedures for inspection by attributes — Part 1: Sampling plans indexed by limited quality level (AQL) for lot-by-lot inspection*. In Table F.2 the procedure for sampling is explained according to this International Standard.

Table F.2 — Procedure for sampling

Process step	Example
Define a sampling method.	Multistage sampling. Selecting enough sampling units so that sample ratio is fulfilled. Sampling is based on weighted features.
Define items.	All features.
Divide the data quality scope (population) into lots.	Number of datasets.
Divide lots into sampling units.	N-number 1 km x 1 km squares.
Define the sampling ratio or the size of the sample.	Sample size depends on the AQL value for that lot.
Select sampling units.	Select required number of sampling units so that sampling ratio or sample size for items is fulfilled.
Inspect items in the sampling units.	Inspect every item in the sampling units.

F.3.2 Sampling methods

If the quality requirements for the feature is one nonconformity per 100 units (Acceptable Quality Level (AQL) = 1), then all features collected are checked from the data source. Inspection by sampling is done when the AQL= 4 or 15. The inspection level is the general inspection level 1, i.e., the single sample program for normal inspection. In the standard there are three general inspection levels (I, II, III) and four additional special levels (S-1,S-2,S-3 and S-4). In general, inspection level defines the sample size from the lot size.

A lot used for testing must consist of datasets produced as far as possible at the same time and with the same methods. From the lot, sampling units of N-number 1 km x 1 km squares are selected so that the number of features in the sample is sufficient for an AQL=4.

Sampling is done using pre-established default weights for the features. In weighting, a default value of 1 is given to features of which there are many in the lot or for which no AQL for completeness has been established. Features whose completeness AQL=4 or 15 are given a weight of 2 or 3. A weight of 3 is given to features that are sparse in the lot. Otherwise a weight of 2 is used. A procedure is available for situations where the required sample size is not achieved. In general the program tries to use special inspection levels S-1 to S-4.

All features in the sampled squares are checked in the field. A feature is non-conforming if it is missing (omission) or if the feature in the dataset does not exist in the field (commission).

F.3.3 Full inspection

A full inspection is carried out on those features which have a quality requirement AQL=1.

F.4 Inspection for quality

A report of the type shown below is completed in the field for each sample test area, and the results are summarised to generate the report shown in Figure F.1. In the sample test area there were, for example, 28

© ISO 2003 — All rights reserved

other buildings of one to two floors, one was missing in the dataset (omission) and there were 11 features that should not have been collected according to product specification.

Name of the Dataset: L213101C Sampling unit: coordinates (North (m),East (m)): 6741000 2509000, length (m), width (m): 1000 1000				
Feature	Number of items	Completeness		Thematic accuracy Classification correctness Number of errors
		Omission Number of errors	Comission Number of errors	
Road class IIa	4			
Road class IIIa	6			
Building, residential, (one to two floor)	10			
Building, other building, (one to two floor)	28	1	11	

Figure F.1 — Example of a field quality check annotations for a portion of an area

F.5 Determination of data quality results and conformance

A computer generated report is produced for each quality test conducted. The full report of quality tests includes over 65 features, some with one or more attributes. Figure F.2 shows an example of a completeness and thematic accuracy evaluation report of the topographic datasets.

In Figure F.2 there are 16 databases (1:10 000 map sheets) which are chosen for the sample. The computer algorithm is used for selection of 1 km x 1 km squares from those databases. An example of one sampling unit is shown in Figure F.2. A printout of that sampling unit is used in the field together with Figure F.1 and every item is checked for completeness and thematic accuracy. The results are then summarised to Figure F.2.

For example the feature Road can have four errors / 100 unit for completeness and four errors /100 unit in classification. In the databases there were 4712 different roads (a road is line between nodes). In the sample there were 184 items. The ISO 2859 requires 80 items for this lot size and inspection level so the minimum requirements are met. The acceptance value for this sample size is 10 so there can be 10 errors in completeness or in classification. In the sample there were only two errors both in completeness and classification so the test is accepted. For reference there is also the acceptance value for AQL=1 (1 error/ 100 unit). This time also this criteria would have been met.

© ISO 2003 — All rights reserved

TEST NAME: 213101_04										
Date: 09.09.1996 15:15:56										
Area: L213101A L213101B L213101C L213101D L213102A L213102B L213102C L213102D L213103A L213103B L213103C L213103D L213104A L213104B L213104C L213104D (Area is defined by map sheets)										
Feature type	Attribute data type	Lot size	Sample size	Size at inspection level 1	Inspection level	AQL	Acceptance value	Acceptance value for AQL 1	Completeness (omission or commission) Number of errors	Thematic accuracy Classification correctness Number of errors
ROAD		4712	184	80	I	4	10	7	2	-
	class (Ia -IIIa)					4	10	7	2	
	road number					1	0	0		
	road section number					1	0	0		
	vertical status					1	0	0		
	one-way traffic					0	184	80		
	pavement type					0	184	80		
	status					0	184	80		
	free height					0	184	0		
BUILDING		6447	222	80	I	4	14	7	4	4
	use					0	222	80	2	
	number of floors					0	222	80		

NOTE Some features are not shown in this example for clarity

Figure F.2 — Completeness and thematic accuracy of topographic database

F.6 Reporting quality results

Figures F.3 and F.4 gives an example how to report the quality results. In Figure F.3 the quality results have been reported for metadata described in ISO 19115. A Quality Evaluation Report is then used to report detailed quality information (Figure F.4). In the parenthesis there is an explanation of used codes that can be found from ISO 19115, but they are not part of the report.

© ISO 2003 — All rights reserved

DataQuality		
dqScope		
scpLvl	**012 (feature type)**	**003 (feature attribute class)**
scpExt	**Extent**	
exDesc	**Lot area**	
geoEle		
exTypeCode	**1 (inclusion)**	
BoundPoly		
polygon	**6740000,2500000,6770000,2500000,6770000,2510000,6760000,2510000,6750000,2510000,6750000,2520000,6740000,2520000,6740000,2500000**	
dqReport		
eleTypCode	**001 (completeness)**	**005 (thematic accuracy)**
subEleCode	**001 (additional) commission and omission**	**002 (classification correctness)**
addSubEle		
addName	**commisson and omission**	
addDesc	**Commisson and omission of the dataset**	
dqResult		
measName	**number of excess or missing item**	**number of errors**
dateTime	**1996-09-09**	**1996-09-09**
measResult		
Result ConResult conSpec		
resTitle	**The Quality model of the TDB**	
resRefDate	**1996**	
conExpl	**conformance to product specification**	
Pass	**1 (pass)**	**1 (pass)**
QuanResult		
quanValDom	**number**	**number**
quanRes	**2**	**2**

Figure F.3 — Reporting as metadata according to ISO 19115

© ISO 2003 — All rights reserved

reportidentification	**Quality Evaluation Report for the Topographic Database**	
Reportscope	**Scope defined in metadata (see: dqScope)**	
compQuantDesc		
dataQualityMeasure		
mathDesc	**Number of missing or excess items**	**Number of errors**
compMeasValue	**2**	**2**
valType	**number**	**number**
realibilityValue	**99**	**99**
realibilityValueUnit	**percent**	**percent**
conformRealibility		
conformRelValues	**AQL=4**	**AQL=4**
conformRelDom	**number**	**number**
ReferenceDoc	**Quality manual of topographic data, Quality model of Topographic data, Instructions for topographic data compilation, Feature catalogue and definitions**	
dqeMethodTypeInfo		
dqeMethodType	**1 (direct external)**	
dqeSamplingApplied	**1 (sampling)**	
dqeMethodInfo		
dqeTheory	**see ISO 2859 and Quality model of Topographic Data**	
dqeProcAlgorithm	**Following program is used in testing: MLAATU.EXE command procedures: LAADUNTARKISTUS.COM and parameters are defined in files: P99.p99,P97.p97, p98.p98**	
dqeParamInfo		
dqeParamDefinition	**Acceptable quality level (see ISO 2859)**	**Acceptable quality level (see ISO 2859)**
dqeParamValues	**4**	**0 (not defined)**
dqeParamDomain	**AQL number**	**AQL number**
dqeParamInfo		
dqeParamDefinition	**Lot size**	**Lot size**
dqeParamValues	**4712**	**6447**
dqeParamDomain	**number**	**number**
dqeParamInfo		
dqeParamDefinition	**Sample size**	**Sample size**
dqeParamValues	**184**	**222**
dqeParamDomain	**number**	**number**
dqeParamInfo		
dqeParamDefinition	**Sample size required at inspection level 1**	
dqeParamValues	**80**	
dqeParamDomain	**number**	
dqeParamInfo		
dqeParamDefinition	**Inspection level**	
dqeParamValues	**I**	
dqeParamDomain	**class**	
dqeParamInfo		
dqeParamDefinition	**Acceptance value**	**Acceptance value**
dqeParamValues	**10**	**222**
dqeParamDomain	**number**	**number**
dqeSampleMethod		
dqeSamplingScheme	**From the lot, an area of so many 1km x 1km squares are sampled that the number of roads in the sample is at least the same as AQL=4 requires.**	
dqeItemDescription	**Item is a road line between nodes**	**Item is a building**
dqeLotDescription	**A lot is group of databases (1:10 000 map sheet) which are taken for inspection. The lot size is number of features in the lot**	
dqeSamplingRatio	**On average an area comprising 4 map sheets (16 databases) with 6 to 10 1 km x 1 km squares is recommended as a practical lot size**	

Figure F.4 — Quality Evaluation Report according to ISO 19114 Annex I

© ISO 2003 — All rights reserved

Annex G
(informative)

Example of measurement and reporting of completeness and thematic accuracy

G.1 Introduction

This annex provides an example of measurement and reporting of thematic accuracy and completeness. The objective of this example is as follows:

— demonstrate how quality evaluation procedures can be applied to measure and report quantitative data quality results,

— provide an example of measurement and reporting of thematic accuracy and completeness,

— demonstrate the use of misclassification matrices as a tool for data quality evaluation.

The example includes how data quality results may be reported in metadata and as a quality evaluation report.

G.2 Dataset description

The 'real world' is represented by Figure G.2. The product specification, given in Figure G.1, describes the universe of discourse. The specification defines those features, attributes, and relationships that are considered important and should be in the dataset.

For the purpose of demonstrating how the dataset may have been produced, the universe of discourse, i.e., the ideal dataset that meets the product specification, is graphically depicted in Figure G.3. In all the figures,

— the digit or letter representing domain of digits under the symbol of a tree is the height of the tree in metres.

— the digit in the symbol of a house is the number of occupants of the house.

— the name of the occupants of a house is noted beside the symbol of the house.

The relationship between the three figures is,

— Figure G.2 represents the 'real world', which generally contains more features than will be contained in the dataset,

— Figure G.3 represents the 'universe of discourse' given by the product specification. It is that part of the 'real world' that is to be included in the dataset, if the dataset is completely and accurately produced,

— Figure G.4 represents the dataset as produced.

© ISO 2003 — All rights reserved

Underlined item is feature type. Listed below each feature type are zero or more attribute names. Each attribute name is followed by a value type of string or integer and separated from the attribute name by a colon. Each value type is followed by an optional value domain enclosed by braces.

Feature types

<u>Industrial building</u>

<u>House</u>
family name : string
number of occupants : integer

<u>Tree</u>
height class : string {A : from 1 to 3 metre, B : from 3 to 5 metre, C : from 5 to 10 metre, D : more than 10 metre}

<u>Path</u>

<u>Road</u>
Condition : string {surfaced, unsurfaced}

Rules from product specification

— trees with a height of less than 1 metre shall not be recorded

— the attribute "condition" of a road may have no value ("undetermined value")

— the attributes "name" and "number of occupants" of a house may have no value ("undetermined value")

Figure G.1 — Product specification

© ISO 2003 — All rights reserved

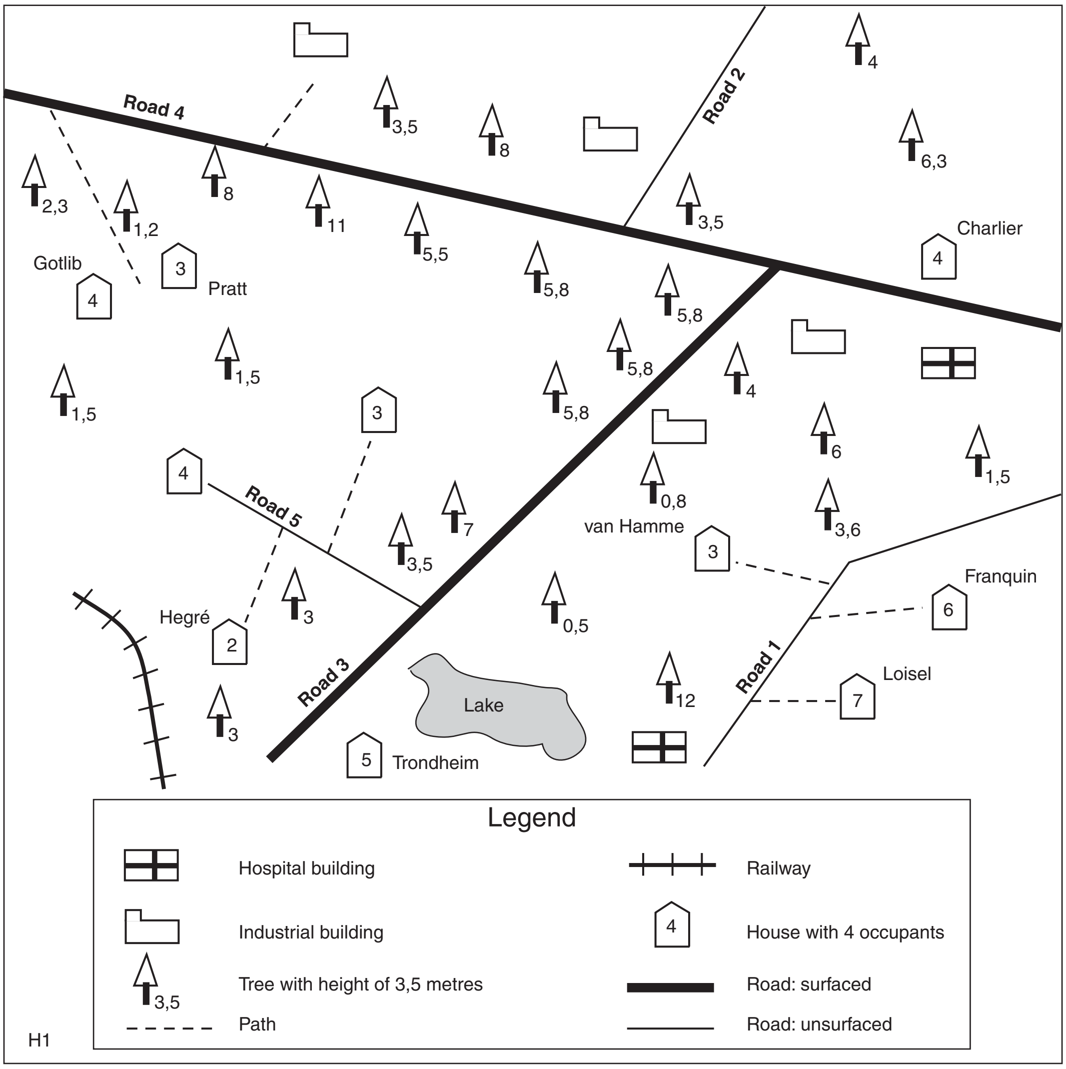


Figure G.2 — The graphical representation of the 'real world'

© ISO 2003 — All rights reserved

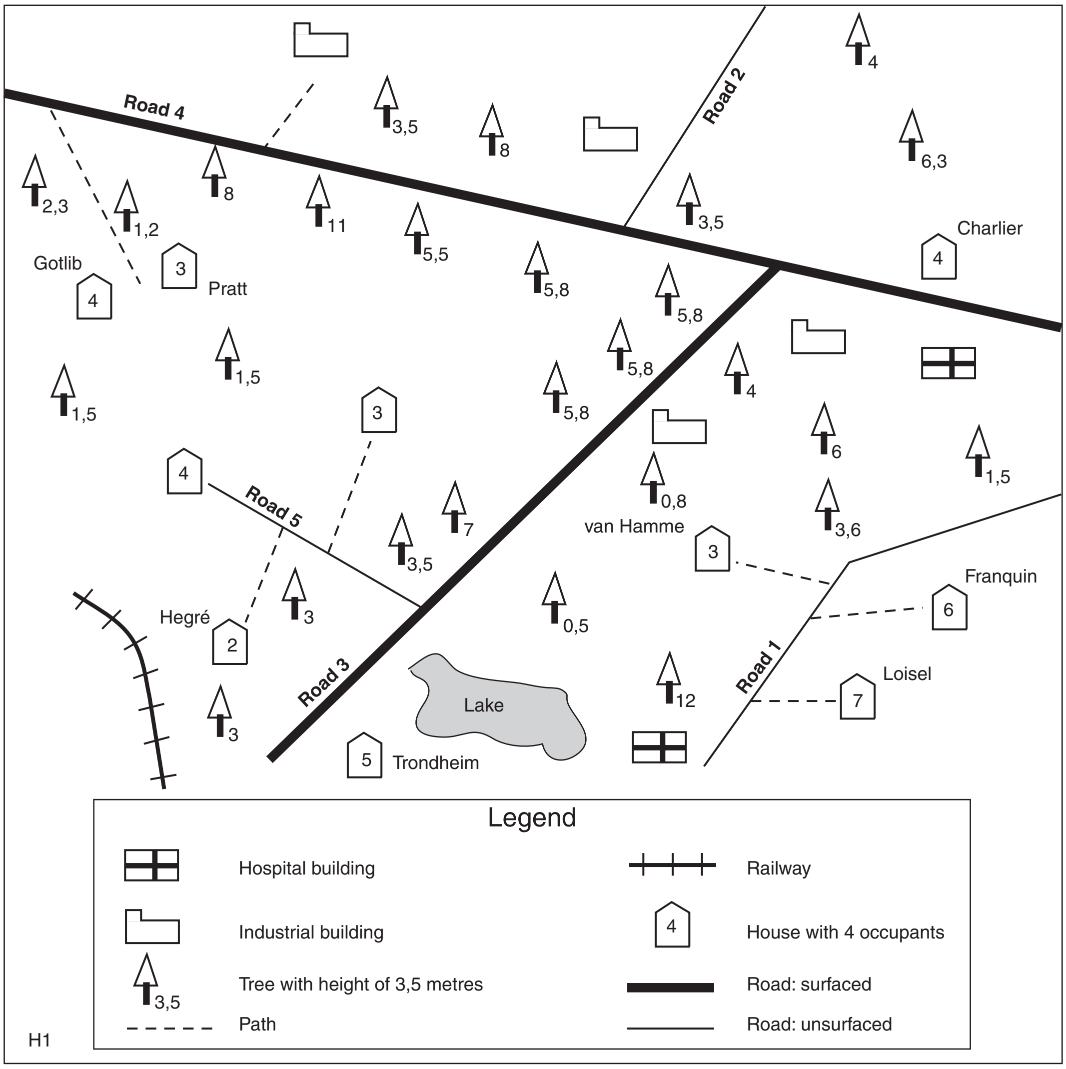


Figure G.3 — Graphical representation of the universe of discourse

© ISO 2003 — All rights reserved

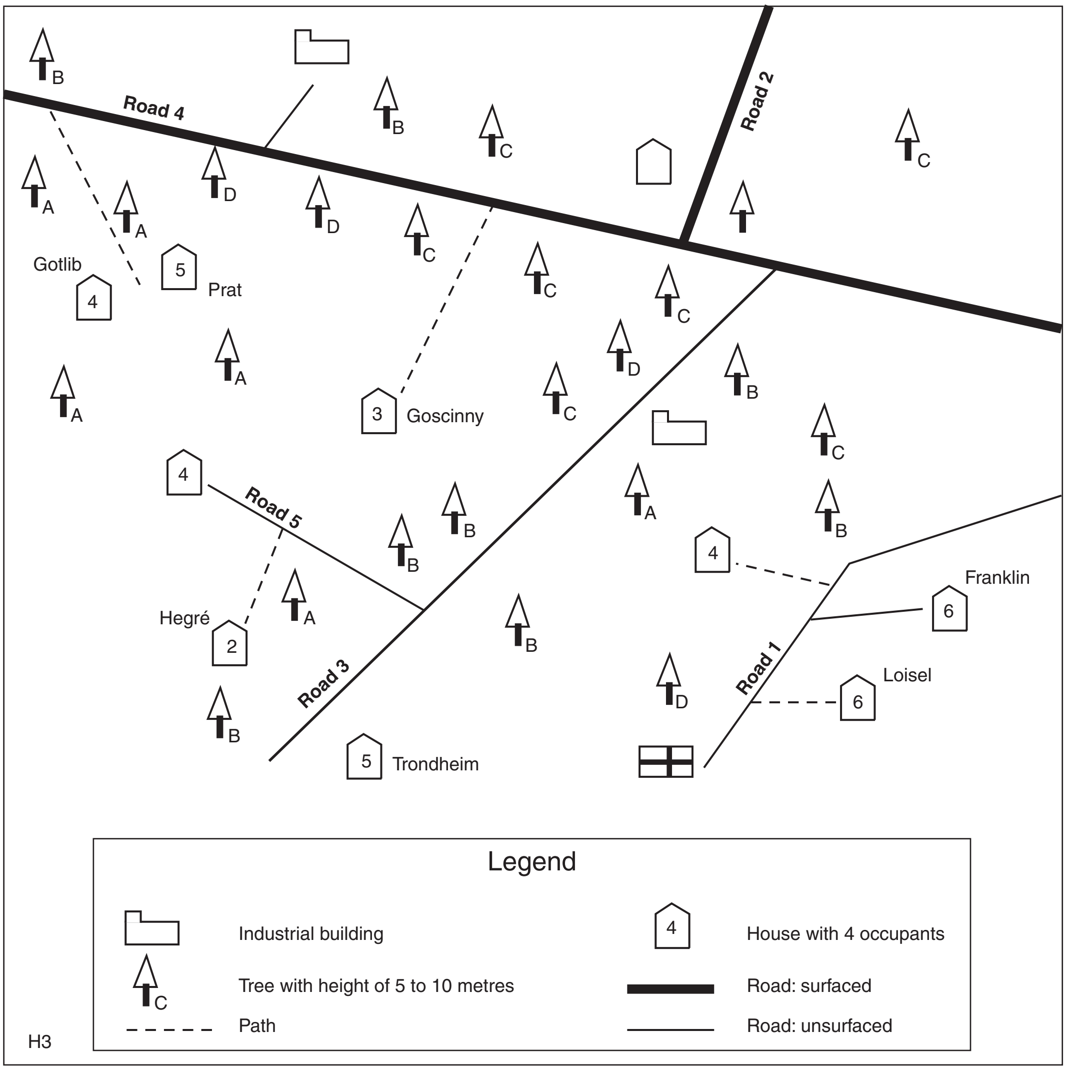


Figure G.4 — Graphical representation of the dataset

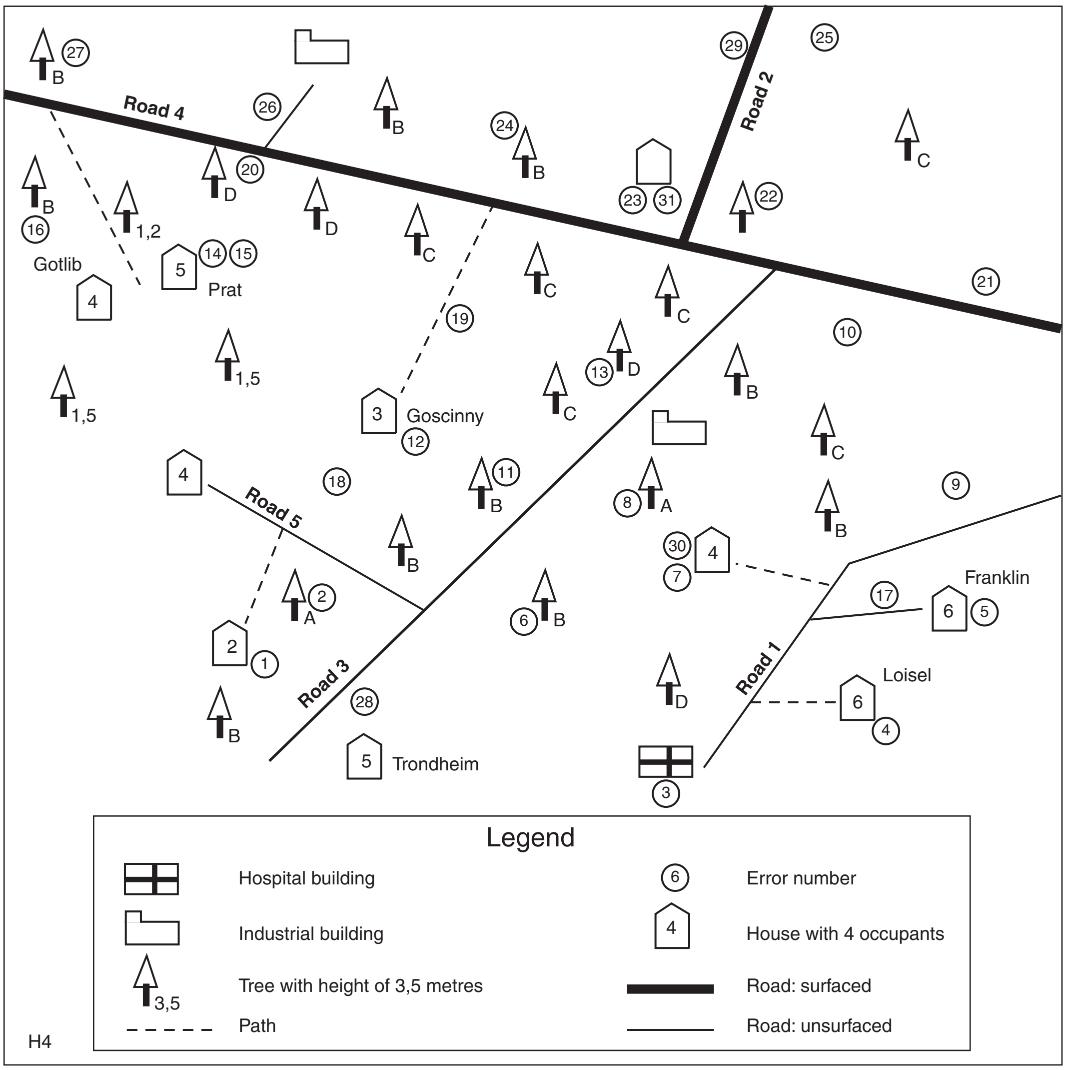


Figure G.5 — Graphical representation of dataset error locations

G.3 Evaluation of data quality

G.3.1 Identification of errors

By comparing the dataset, represented by Figure G.4, with the universe of discourse, represented by Figure G.3, a list of errors in the example dataset can be produced. The following is a list of detected errors with error numbers given for reference.

1) Errors of omission and commission in recording of trees. Three trees (#6, #8, #27) are in excess and two trees are missing (#9, #25).

2) Errors of omission and commission in recording paths. One path is missing (#18) and one is in excess (#19).

3) A house replaces an industrial building (#23).

4) Two paths are miscoded as roads (#17, #26).

 © ISO 2003 — All rights reserved

5) A house is missing (#21).

6) Attribute error on roads. Two roads have the wrong ‘condition’ (#29, #28).

7) A hospital is represented in the dataset (#3).

8) Two trees with a height less than 1 m are represented in the dataset (#6, #8)

9) Tree height attribute class code missing. A tree is missing a class code while it is ‘B’ in the universe of discourse (#22).

10) Tree height attribute misclassified. Six trees have the wrong height class assigned (#2, #11, #13, #16, #20, #24).

11) House name attribute ‘family name’ errors. The houses named “van Hamme” (#7) and “Hergé” (#1) in the universe of discourse have no name in the dataset. The house named “Goscinny” in the dataset (#12) has no name in the universe of discourse.

12) House name attribute ‘family name’ errors. The houses named “Franquin” (#5) and “Pratt” (#15) in the universe of discourse are named “Franklin” and “Prat” respectively in the dataset.

13) House occupant count attribute errors. The occupant count attribute is missing for one house (#31) and wrong for three houses (#4, #14, #30).

14) Omission error in industrial buildings. One industrial building is missing (#10).

NOTE The classification of errors as omission/commission, completeness or thematic accuracy is subjective. For example, the misclassification of a house as an industrial building could alternately be considered as an error of omission of the one and commission of the other.

G.3.2 Completeness

ISO 19113 defines completeness as the presence and absence of features, their attributes and their relationships. Completeness in this example is classified by feature class. The types of measures tested for are commission and omission. Table G.1 depicts a way to classify completeness.

Table G.1 — Completeness by feature class

Feature class	Number of instances in the universe of discourse	Commission – count	Commission – percentage$_1$	Omission – count	Omission – percentage$_2$
Path	7	0	0	2	29
Road	5	2	40	0	0
Tree	25	3	12	2	7
Industrial building	4	0	0	2	50
House	10	1	10	1	10
Hospital	0	1	100	0	0

NOTE 1 Commission percentage = number of included items/number of items in the universe of discourse * 100

NOTE 2 Omission percentage = number of omitted items/number of items in the universe of discourse * 100

© ISO 2003 — All rights reserved

G.3.3 Thematic accuracy

ISO 19113 defines thematic accuracy as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. One way of depicting errors associated with thematic accuracy is by using a “misclassification matrix".

NOTE 1 A misclassification matrix is a square matrix where the i,j element corresponds to the quantity classified as belonging to class j when it actually belong to class i.

Table G.2 is a misclassification matrix that shows errors by feature class. It explains how well the instances in the dataset are classified. The different percentages should always refer to the population in the dataset.

NOTE 2 In the matrices, the number after the name of the feature type denotes the number of occurrences and the value in the cell is the percent of misclassification.

Table G.2 — Feature misclassification matrix

universe of discourse	**dataset**					
	Path 5	**Road 7**	**Tree 26**	**Industrial building 2**	**House 10**	**None (omitted features)**
Path 7	4 / 5 = 80%	2 / 7 = 29%	0%	0%	0%	1 / 5 = 20%
Road 5	0%	5 / 7 = 71%	0%	0%	0%	0%
Tree 25	0%	0%	23 / 26 = 88%	0%	0%	2 / 5 = 40%
Industrial building 4	0%	0%	0%	2 / 2 = 100%	1 / 10 = 10%	1 / 5 = 20%
House 10	0%	0%	0%	0%	9 / 10 = 90%	1 / 5 = 20%
None 3 (committed features)	0%	0%	3 / 26 = 11%	0%	0%	
Sum dataset population	5 / 5 = 100%	7 / 7 = 100%	26 / 26 = 100%	2 / 2 = 100%	10 / 10 = 100%	

In tables G.3 and G.4, only features that have homologue in the same feature type ("class") are taken into account.

— Attribute height of trees

© ISO 2003 — All rights reserved

Table G.3 — Feature attribute height misclassification matrix – tree height

universe of discourse	dataset				
	class A 1 to 3 m 5	**class B 3 to 5 m 10**	**class C 5 to 10 m 5**	**class D > 10 m 4**	**in-determined (missing values) 4**
class A 5	3 / 5 = 60%	1 / 10 = 10%	0%	0%	1 / 4 = 25%
class B 8	1 / 5 = 20%	5 / 10 = 50%	0%	0%	2 / 4 = 50%%
class C 10	0%	2 / 10 = 20%	5 / 5 = 100%	2 / 4 = 50%	1 / 4 = 25%
class D 2	0%	0%	0%	2 / 4 = 50%	0%
in-determined 3 (commission)	1 / 5 = 20%	2 / 10 = 20%	0%	0%	0%
sum dataset population	5 / 5 = 100%	10 / 10 = 100%	5 /5 = 100%	4 / 4 = 100%	4 / 4 = 100%

— Attribute condition of roads

Table G.4 — Feature attribute misclassification matrix – road condition

universe of discourse	dataset	
	surfaced 2	**unsurfaced 3**
surfaced 2	1 / 2 = 50%	1 / 3 = 33%
unsurfaced 3	1 / 2 = 50%	2 / 3 = 67%

— Attribute "number of occupants" of houses as an example of accuracy of a quantitative feature attribute defined by a value. The following demonstrates a way to measure the data quality elements thematic accuracy and completeness, and how to express the results of the measurements in terms of text, commission/omission ratios and error statistics.

  — 1/9 houses has no value for the number of occupants

  — bias : -2/8=-0,25 occupants

  — RMSE : 0,87 occupant

  — sample size : 8

G.4 Reporting quality results

G.4.1 Example of error of commission

The following gives an example how to report the quality results for one type of error, commission errors for feature type “path”. First the quality results have been reported as metadata. A data quality evaluation report is then used to report detailed quality information.

G.4.2 Reporting in metadata

Figure G.6 is an example of how to report the quality results as metadata as described in ISO 19115. The explanation of the codes used from ISO 19115 are given in parenthesis, but are not part of the report.

© ISO 2003 — All rights reserved

Figure G.6 — Reporting as metadata according to ISO 19115

G.4.3 Reporting as quality evaluation report

Figure G.7 is an example of how to report the quality results as a data quality report.

addQualityReport	
reportIdentification	Quality Report of Example in this annex
reportScope	Dataset
compQuantDesc	
dataQualMeasure	
mathDesc	Number of items in dataset divided by number of items in universe of discourse multiplied by 100
compMeasValue	ratio
valType	real
realibilityValue	100
realibilityValueUnits	
conformConfidence	
conformConfValue	
conformConfValDesc	
referenceDoc	
dqeMethodTypeInfo	
dqeMethodType	2 *(direct internal)*
dqeSamplingApplies	3 *(not applicable)*
dqeMethodInfo	
dqeAssumptions	
dqeProcAlgorithm	Compare visual count of trees in source with dataset
dqeParamInfo	
dqeParamDefinition	
dqeParamValues	
dqeParamDomain	
dqeFullInspectMetho	
dqeFullInspecType	Count of trees
dqeItemDesc	Trees per product specification
referenceDoc	
dqeSampleMethod	
dqeSamplingScheme	
dqeItemDescription	
dqeLotDescription	
dqeSamplingRation	
dqeDeductiveSource	
dqeDeductRefDocs	
referenceDoc	
aggSourceValues	
aggResult	
aggValueDomain	
aggMeasureValue	
aggErrorStat	
aggQEPreport	
qepOtherDesc	

Figure G.7 — Quality Evaluation Report according to ISO 19114 Annex I

© ISO 2003 — All rights reserved

Annex H
(informative)

Example of an aggregated data quality result

H.1 Introduction

The information in this example is based upon techniques in use in private industry in Europe, North America and Asia. The objective of the example described is to illustrate the techniques of the measurement and aggregation of thematic accuracy, completeness and positional accuracy in a road-based dataset.

This example is concerned only with reporting an aggregated data quality result. No comparison with a conformance quality level is made.

H.2 Dataset description

H.2.1 Real world representation

The real world is represented by Figure H.1. Figure H.1 also depicts a lot drawn from the full dataset of road-based data. The shaded rectangular area at grid square B-2 represents the randomly selected sampling unit to be tested.

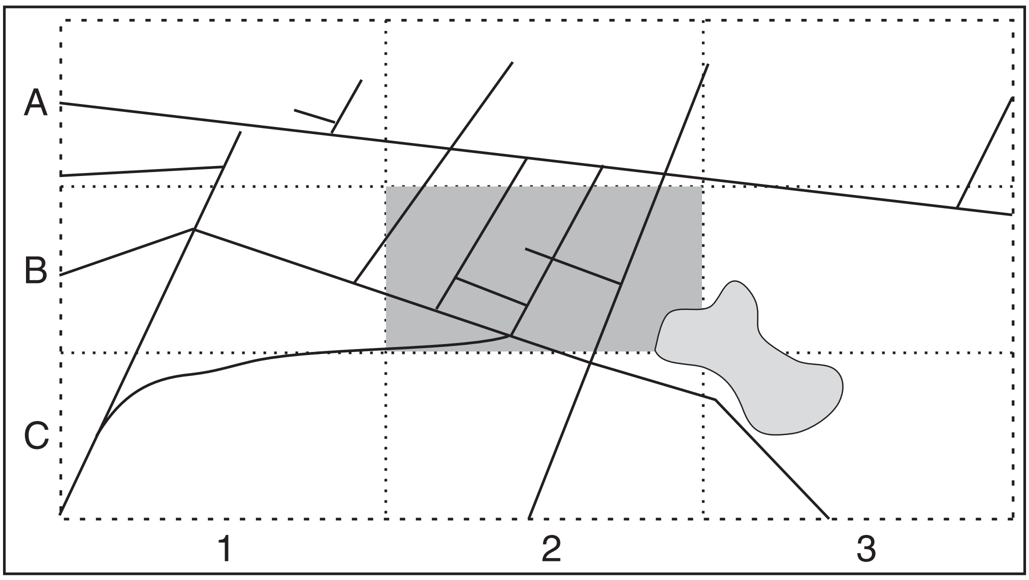


Figure H.1 — Randomly selected lot from full data base and randomly selected sampling unit (darker shaded rectangle)

H.2.2 Product specification

Although abbreviated for the purpose of this example, the product specification defining the universe of discourse is given in Figure H.2. The specification describes those rules that are considered important to the product.

© ISO 2003 — All rights reserved

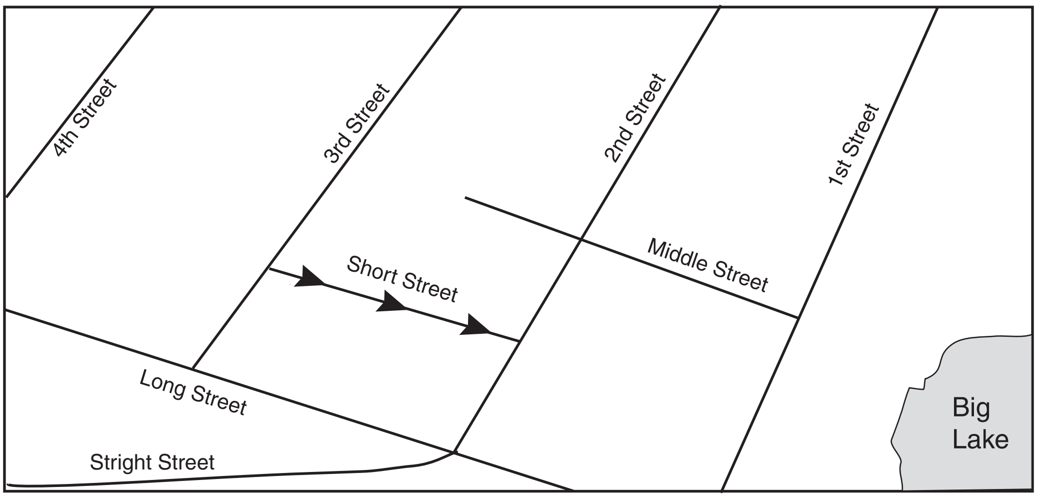
Rules from product specification

- All roads should be included.
- All roads should be named.
- Direction of flow of all one way streets shall be indicated.
- All hydrographic features should be included.

Figure H.2 — Product specification

H.3 Universe of discourse

The universe of discourse is represented in Figure H.3. For purposes of this example, this provides a graphic reference of the reality against which the contents of the dataset will be compared.



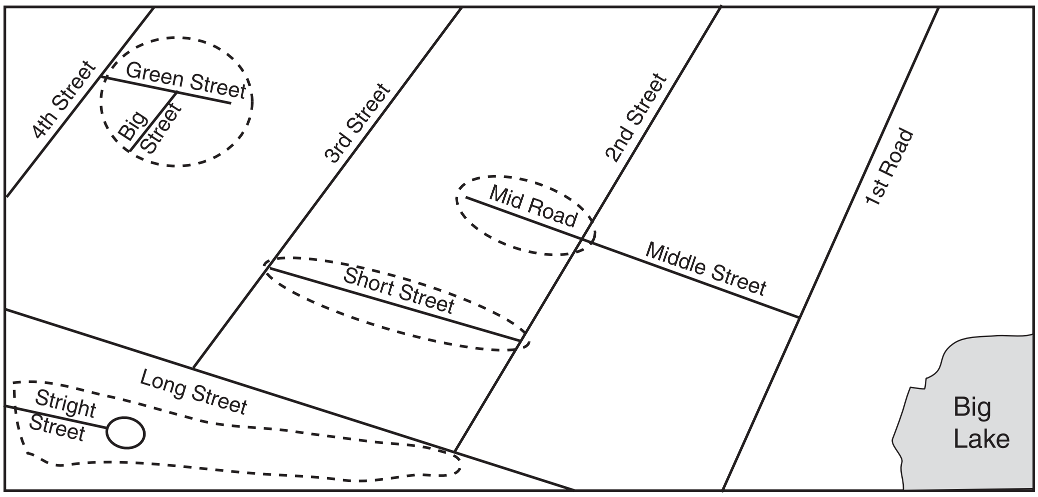
NOTE Arrow indicates direction of traffic flow; no arrow indicates two- way traffic flow.

Figure H.3 — Graphic representation of the universe of discourse

H.4 Dataset

The content of the dataset is represented in Figure H.4. The dotted lines indicate places where errors were detected, i.e., the dataset did not agree with reality. Several types of errors are noted here. Table H.1 identifies the errors and their types.

 © ISO 2003 — All rights reserved



NOTE Arrow indicates direction of traffic flow; no arrow indicates two-way traffic flow.

Figure H.4 — Graphic representation of the dataset contents

Table H.1 — Error types detected and typical data quality subelements under which quality results may be reported

Error types detected	Element	Subelement under which error is reported
Roads that do not exist, e.g., Green Street	Completeness	Commission
Incorrect road names, e.g., 1st Road	Thematic accuracy	Qualitative attribute correctness
Missing part of road, e.g., Straight Street	Logical consistency	Topological consistency
Missing attribute data, e.g., Short Street flow arrow	Thematic accuracy	Qualitative attribute correctness[a]

[a] If the rules for the database given in the product specification requires the flow direction field to always have an entry, such as one-way or two way traffic flow, the error is measured as an omission, however, if only an entry is required, it is measured as thematic correctness.)

H.5 Aggregation of evaluation results and reporting

An error table is prepared to show the number of errors encountered and how they are classified according to a typical procedure used in the road database industry. The particular example procedure assigns weights to each error type. The sum of the weights equals 100 percent. The resulting weighted value is considered to be the quality of the dataset. Table H.2 shows an example of calculating an aggregated data quality result. An item is defined as a road segment which is bounded by intersection points with the other roads or boundaries of sample unit.

© ISO 2003 — All rights reserved

Table H.2 — Example of computation of an aggregated quality evaluation result

Feature	Number of items in lot	Number of non-conforming items	Ratio of non-conforming	Accuracy Proportion (defined as 1-ratio)	Weights	Weighted value (accuracy proportion * weight)
Road segment	19					
Incorrect		1				
Missing		0	4 / 19	0,79	50%	0,39
Excess		3				
Street Name						
Base name	19	5	5 / 19	0,74	15%	0,11
Direction-of-travel	19	1	1 / 19	0,95	25%	0,23
Hydrography	1	0	0 / 1	1,00	10%	0,10
Aggregated data quality result (defined as sum of weighted accuracy proportion * 100)						84%

© ISO 2003 — All rights reserved

Annex I
(normative)

Reporting quality information in a quality evaluation report

I.1 Introduction

This annex describes the content of a detailed quantitative quality evaluation report. The quality evaluation report provides more detail about the quality results and the procedures used to compute them than is recorded in metadata. Table I.1 of this annex provides a graphic of the nested relationships of the quality evaluation report content.

I.2 Quality evaluation report components

The table column heading and table codes in Table I.1 are:

Table line number provides a reference for each item in the table and is used in domain column to show range of this item's components in the table.

Name report element name.

Definition / content defines the item or describes the content of the item.

Obligation / condition gives requirements for reporting the item or the conditions under which the item is required. There are three obligation codes:

Mandatory (M) denotes an is required entry.

Conditional (C) entry required when the stated condition is satisfied.

Optional (O) entry is optional.

Maximum occurrences (max occur) maximum times this item can occur within a superior item's domain. An integer entry indicates that number of times, and N indicates as many as desired.

Data type report section, text, entity, or when not applicable, a "–" is shown.

Domain for each report element, the domain specifies the values allowed or the use of free text. 'Free text' indicates that no restrictions are placed on the content of the entry. Integer-based codes shall be used to represent values in restricted (closed) domains.

© ISO 2003 — All rights reserved

Table I.1 — Quality evaluation report components

Line #	Name	Definition / content	Obligation / condition	Max occur	Data type	Domain
1	addQualityReport	Quality evaluation report	C / subclause 9.2	1	report section	Lines 2 - 40
2	reportIdentification	Report identification information	M	1	CharacterString	Free text
3	reportScope	Scope of dataset evaluated in this report (ISO 19113)	O	1	CharacterString	MD_MetadataScope <<CodeList>>
4	compQuantDesc	Complementary description of quantitative assessment such as data quality measure values and their reliability limits	M	1	report section	Lines 5 - 14
5	dataQualMeasure	Information on definition and value of data quality measure of an object data quality scope	M	1	report section	Lines 6 - 10
6	mathDesc	Mathematical description of data quality measure	M	1	CharacterString	Free text
7	compMeasValue	Values of data quality measure applied	M	1	CharacterString	Free text
8	valType	Unit in which data quality measure value is recorded	M	1	CharacterString	Free text
9	realibilityValue	Reliability or confidence limit valuess of the computed or estimated data quality measure value	O	1	CharacterString	Free text
10	realibilityValueUnits	Unit in which reliability values are recorded	O	1	CharacterString	Free text
11	conformConfidence	Confidence in conformance	O	1	report section	Lines 12 - 14
12	conformConfValue	Confidence in the conformance result. NOTE The confidence in the conformance may be such as HIGH,LOW,NONE, or 95%, or so forth	M	1	CharacterString	Free text
13	conformConfValDesc	Unit or value type in which the confidence in conformance is recorded	M	1	CharacterString	ValueUnit or ValueType
14	referenceDoc	Information on documents which are referenced in developing and applying the data quality evaluation method	O	N	Class	CI_Citation
15	dqeMethodTypeInfo	Detailed information about applying the quality evaluation method	M	1	report section	Lines 16 – 37
16	dqeMethodType	Quality evaluation method class	M	1	CharacterString	1 - direct-external 2 - direct-internal 3 - indirect
17	dqeSamplingApplied	Information on inspection strategy applied	M	1	CharacterString	1 – sampling applied 2 – full inspection 3 – not applicable
18	dqeMethodInfo	Information on the data quality evaluation method	M	1	report section	Lines 19 - 37

© ISO 2003 — All rights reserved

Table I.1 *(continued)*

Line #	Name	Definition / content	Obligation / condition	Max occur	Data type	Domain
19	dqeAssumptions	Information on underlying assumptions in developing and applying the data quality evaluation method	O	1	CharacterString	Free text
21	dqeProcAlgorithm	Information on how data is processed to determine the data quality result	M	1	CharacterString	Free text (if a specific computer algorithm or command is used, then its name shall be included)
22	dqeParamInfo	Information on parameters used in the data quality evaluation method	O	N	report section	Lines 23 - 37
23	dqeParamDefinition	Information on the definition of parameter used	M	1	CharacterString	Free text, e.g., weight value of each aggregate data quality measure
24	dqeParamValues	Value of parameter used in the data quality evaluation method	M	1	CharacterString	Free text
25	dqeParamDomain	Unit in which the parameter value is recorded	M	1	CharacterString	Free text
26	dqeFullInspecMethod	Information on full inspection method	C / full inspection applied	1	Report section	Lines 27 - 29
27	dqeFullInspecType	Information on the type of full inspection and description of the procedure	M	1	CharacterString	Free text
28	dqeItemDescription	Information on how items are defined	M	1	CharacterString	Free text
29	referenceDoc	Information on documents which are referenced in developing, applying the data quality evaluation method	O	N	Class	CI_Citation
30	dqeSampleMethod	Information on sampling method	C / sampling applied	1	report section	Lines 31 - 37
31	dqeSamplingScheme	Information on the type of sampling scheme and description of the sampling procedure	M	1	CharacterString	Free text, e.g., simple random sampling: items are sampled from each lot.
32	dqeItemDescription	Information on how items are defined	M	1	CharacterString	Free text
33	dqeLotDescription	Information on how lots are defined	C / lot applied	1	CharacterString	Free text
34	dqeSamplingRatio	Information on how many samples on average are extracted for inspection from each lot or population	M	1	CharacterString	Free text

© ISO 2003 — All rights reserved

Table I.1 *(continued)*

Line #	Name	Definition / content	Obligation / condition	Max occur	Data type	Domain
35	dqeDeductiveSource	Information on what data are used as sources in deductive evaluation method	C / deductive method applied	1	CharacterString	Free text, e.g. lineage and usage of the data quality scope
36	dqeDeductRefDocs	Identification of source documents used as basis for deduction	M	N	CharacterString	Free text
37	referenceDoc	Information on documents which are referenced in developing and applying the data quality evaluation method	O	N	Class	CI_Citation
38	aggSourceValues	Information on which component datasets are used and what data quality measures are aggregated for determining the data quality measure value and conformance	C / aggregation result computed	N	report section	Lines 39 - 44
39	aggResult	Description of the value as a quantitative result	M	1	report section	Lines 40 - 44
40	aggValueDomain	Unit in which the quantitative value is recorded	M	1	CharacterString	Free text, e.g., metres, kilometers
41	aggMeasureValue	Value of measure applied	M	1	CharacterString	Free text
42	aggErrorStat	Type of the statistic	M	1	CharacterString	Free text, e.g., RMS
43	dateTime	Data and time when the value was computed	O	1	DateTime	ISO 19108
44	aggQEPreport	A pointer to an quality evaluation report	O	1	Class	CI_Citation
45	qepOtherDesc	Additional information, including intermediate results, that is considered important when estimating data quality measure values and determining conformance	O	N	CharacterString	Free text

© ISO 2003 — All rights reserved

Annex J
(informative)

Aggregation of data quality results

J.1 Introduction

Quality of a dataset may be represented by one or more aggregated data quality results (ADQR). The ADQR combines quality results from data quality evaluations based on different data quality elements, data quality subelements and/or data quality scopes.

The following subclauses of this annex are examples of methods that may be used for producing an ADQR. While the examples show computation using Boolean values, they do not have to be Boolean. A data quality result may be quantitative or qualitative and represented by a numeric or Boolean value. A dataset may be deemed to be of an acceptable aggregate quality even though one or more individual data quality results fails acceptance. In any case, the meaning of the aggregate result must be made clear.

As the ADQR may be difficult to fully understand, the meaning of the aggregate data quality result should be understood before drawing conclusions based on aggregate data quality results for the quality of the dataset.

Clause 8 of this International Standard describes reporting requirements for aggregate data quality results.

J.2 100% pass/fail

Each data quality result involved in the computation is given a Boolean value v of one (1) if it passed and zero (0) if it failed. The aggregate quality is determined by the equation,

$$ADQ = v_1 * v_2 * v_3 * . . . * v_n$$

where n is the number of data quality measurement frames.

If ADQR = 1, then the overall dataset quality is deemed to be fully conforming, hence pass. If ADQR = 0, then it is deemed non-conforming, hence fail. The technique does not provide a result that indicates location or magnitude of the non-conformance.

J.3 Weighted pass/fail

Each data quality result involved in the computation is given a Boolean value v of one (1) if it passed and a zero (0) if it failed. Based on the significance to the purpose of the product, a weight value w between 0,0 and 1,0, inclusive, is assigned to each data quality result. The total of all the weights should equal 1,0. The choice of weights is a subjective decision made by the data producer or user. The reason for the data producer's decision should be reported as part of the result. The aggregate quality is determined by the equation,

$$ADQR = v_1 * w_1 + v_2 * w_2 + v_3 * w_3 + . . . + v_n * w_n$$

where n is the number of data quality measurement frames.

The technique does provide a magnitude value indicating how close a dataset is to full conformance as measured. The technique does not provide a quantitative value that indicates where conformance or non-conformance occurs.

© ISO 2003 — All rights reserved

J.4 Subset of results sufficient for product purpose

This technique is a modification of the 100% pass/fail and the weighted pass/fail methods. A subset of data quality results involved in the computation is selected from data quality results produced during the full data quality evaluation. The subset represents data quality results considered significant to the purpose of the product. This technique may be used when more data quality elements have been measured than are needed to meet the product specification and/or purpose.

The aggregate quality is determined by applying the 100% pass/fail, the weighted pass/fail, or some other aggregate evaluation technique to the subset of data quality measurement frame results.

When this technique is applied, the identity of the data quality measurement frames selected as members of the subset should be documented.

J.5 Maximum/minimum value

Each data quality result is given a value v based on the significance of a data quality result to the purpose of the product. The reason for the data producer’s decision should be reported as part of the dataset’s quality result. The aggregate quality is determined by either of the two equations,

$$ADQR = MAX(v_i , in = 1 . . . n)$$

or

$$ADQR = MAX(v_i , in = 1 . . . n)$$

where n is the number of data quality measurement frames measured.

The technique does provides a magnitude value indicating how close a dataset is to full conformance as measured, but only in terms of the data quality measurement frame represented by the maximum or minimum. The technique does provide a quantitative value that indicates where conformance or non-conformance occurs when the selected data quality measurement frame is reported along with the ADQR. However, this type of ADQR tells little about the magnitude of the other data quality results.

© ISO 2003 — All rights reserved

Bibliography

[1] ISO 2859 (all parts), *Sampling procedures for inspection by attributes*

[2] ISO 3534-2:1993, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control*

[3] ISO 3951:1989, *Sampling procedures and charts for inspection by variables for percent nonconforming*

[4] ISO 8601:2000, *Data elements and interchange formats — Information interchange — Representation of dates and times*

[5] ISO 9000 (all parts), *Quality management and quality assurance standard*

[6] ISO 11404:1996, *Information technology — Programming languages, their environments and system software interfaces — Language-independent datatypes*

[7] ISO 19108:2002 , *Geographic information — Temporal schema*

© ISO 2003 — All rights reserved