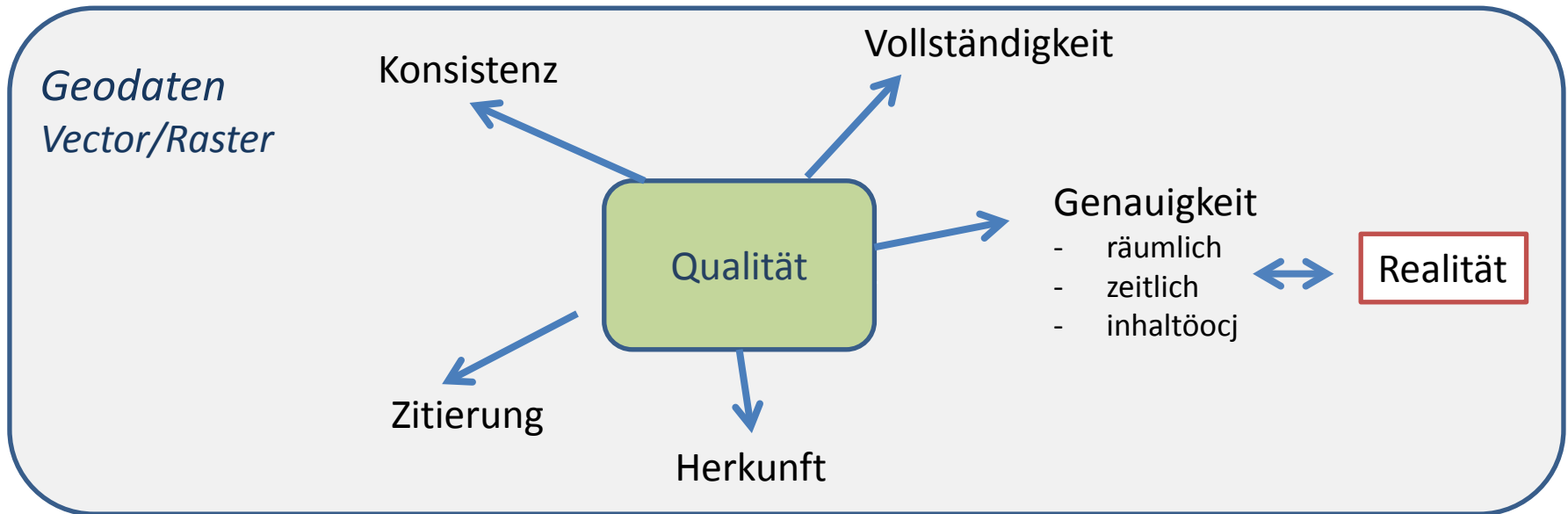


# Datenqualität

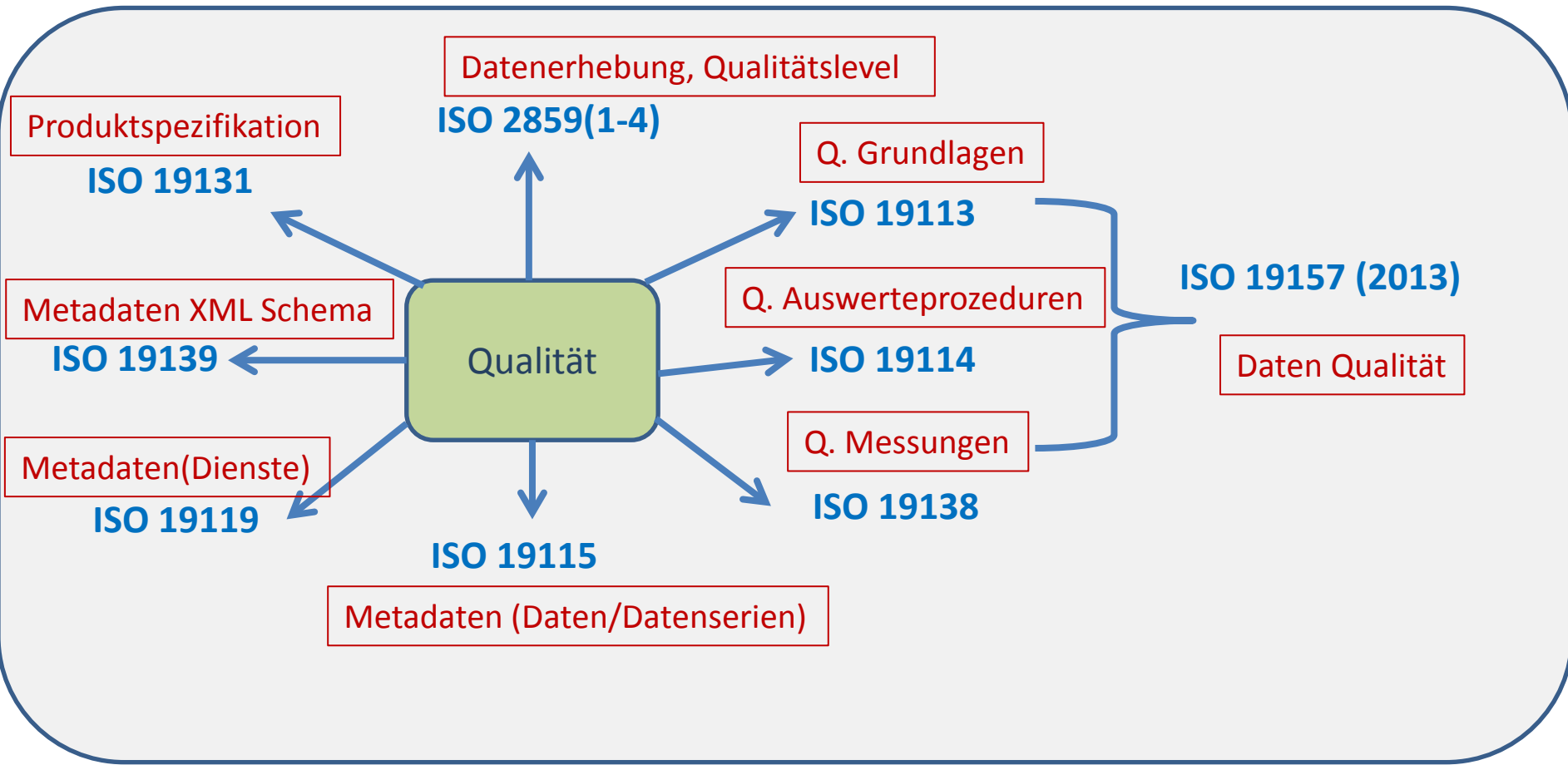
## Komponenten



### Komponenten der Datenqualität (Erzeuger)

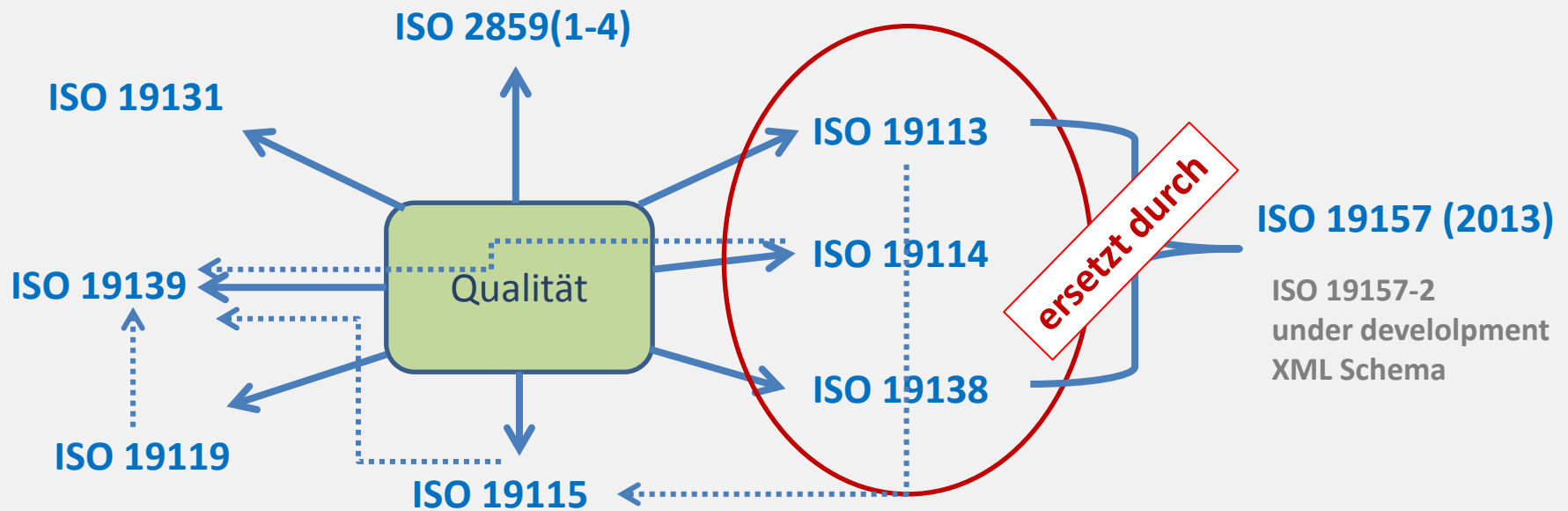
1. **Vollständigkeit:** Elemente vorhanden/nicht vorhanden, Attribute und Relationen;
2. **Logische Konsistenz:** Grad der Genauigkeit bezüglich der logischen Datenstruktur;
3. **räumliche Genauigkeit:** Positionsgenauigkeit;
4. **zeitliche Genauigkeit:** Genauigkeit der zeitlichen Attributwerte;
5. **inhaltliche Genauigkeit:** Genauigkeit quantitativer/nicht quantitative Angaben, Richtigkeit der Klassifikation, Genauigkeit nominaler und ordinaler Daten
6. **Herkunft:** Informationen über die Herkunft der Daten, Vorverarbeitungsschritte, stat. Auswertung, Interpolationen, usw.

# ISO Standards der Datenqualität (I)



# ISO Standards der Datenqualität (II)

ISO 19100 Standard für geographische Informationen



# ISO Metadata related standards

Name of Standard	Description	Status
<b>ISO 19109:2005 Rules for application schema</b>	<b>Defines the a general feature model and rules for creating and documenting application schemas for modeling features and their properties allowing physical applications to understand and share data</b>	<b>Under revision: Sent to ISO for publication</b>
<b>ISO 19110:2005 + Amendment 1 Methodology for feature cataloguing</b>	<b>Catalogue defining features and properties for a domain of interest and/or a dataset and a schema for encoding in XML</b>	<b>Under revision: FDIS 2014-7, IS 2016-02</b>
ISO 19111:2007 Spatial referencing by coordinates	Metadata about/defining a coordinate reference system	International standard
ISO 19111-2:2009 Spatial referencing by coordinates-Extension for parametric value	Metadata about/defining a coordinate reference system using parametric values	International standard
ISO 19112:2003 Spatial referencing by geographic identifiers	Metadata about/defining a reference system which uses spatial unit identifiers other than coordinates i.e. gazetteer, postal code, etc.	International standard
ISO 19113:2002 Quality principles	Defines the principles, the elements/sub-elements of data quality	Superseded by 19157
ISO 19114:2003 Quality evaluation procedures	Defines procedures for determining data quality	Superseded by 19157

<https://www.fgdc.gov/participation/working-groups-subcommittees/mwg/meetings/isostandardsupdate>

# ISO Metadata related standards

Name of Standard	Description	Status
ISO 19115:2003 Geospatial metadata	Defines metadata elements and schema describing geospatial datasets	Superseded by 19115-1
ISO 19115-1:2014 Geospatial metadata fundamentals	Revision of ISO19115 which defines metadata elements and schema describing geospatial resources i.e. datasets and services	International standard
ISO 19115-2:2009 Extensions for imagery and gridded data	Defines additional metadata elements and schema describing imagery and gridded geospatial datasets	Under revision: 1 <sup>st</sup> meeting 2015-06, CD 2016-03, DIS 2016-11, TS 2017-11
ISO 19115-3: XML schema implementation metadata fundamentals	Provides a schema for implementing ISO 19115-1 in XML	Under development: TS 2015-09?
ISO 19119:2005 Services	Provides a framework and defines the metadata for services enabling users to access and process geographic information across a generic computing interface. The metadata portion of this standard has been moved to ISO19115-1	Under revision, Partially superseded by 19115-1: IS 2015-08?
ISO 19130:2010 Imagery sensor models for geopositioning	Specifies a sensor model describing the physical and geometrical properties of specific sensors	International standard

<https://www.fgdc.gov/participation/working-groups-subcommittees/mwg/meetings/isostandardsupdate>



# ISO Metadata related standards

Name of Standard	Description	Status
<b>ISO 19130-1 Imagery sensor models for geopositioning</b>	<b>Revision of ISO 19130</b>	<b>Under revision: CD 2015-12 DIS 2016-06, IS 2017-06</b>
ISO 19130-2:2014 Imagery sensor models for geopositioning SAR, InSAR, Lidar and Sonar	Specifies a sensor model describing the physical and geometrical properties for the stated sensors	International standard
ISO 19138:2006 Data quality measures	Defines commonly used measures for reporting data quality for the sub-elements defined in ISO 19113 and a structure so they may be maintained in a register.	Superseded by 19157
ISO 19139:2007 Metadata XML Schema implementation	Provides encoding rules and a schema for implementing ISO 19115 in XML.	International standard To be partially superseded by ISO 19115-3
<b>ISO 19139-1 Metadata XML Schema implementation</b>	<b>This revision will only include the encoding rules for metadata</b>	<b>Under revision: CD 2015-06 DIS 2016-06, TS 2017-06</b>
ISO 19139-2:2012 Metadata - XML schema for imagery and gridded data	Provides a schema for implementing ISO 19115-2:2009 in XML	International standard

<https://www.fgdc.gov/participation/working-groups-subcommittees/mwg/meetings/isostandardsupdate>

# ISO and W3C Metadata related standards

Name of Standard	Description	Status
ISO 19157:2013 Data Quality	Defines the principles and components for describing, evaluating, and the measures used for reporting data quality. Revising and replacing ISO 19113, 19114, 19138.	International standard
<b>ISO 19157-2 Data Quality XML Schema implementation</b>	<b>Provides a schema for implementing ISO 19157 in XML</b>	<b>Under development: TS 2016-05</b>
ISO 15836:2009 The Dublin Core metadata element set	Cross domain resource descriptions – not limited to specific resources	International standard Reviewed 2014

Name of Standard	Description	Status
W3C Data Catalog Vocabulary (DCAT)	An RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.	International standard

<https://www.fgdc.gov/participation/working-groups-subcommittees/mwg/meetings/isostandardsupdate>

# EN ISO 19113

veröffentlicht 2002 - geht über in -> ISO 19157 (2013)

- Beschreibung der Qualität von Geodaten
- Spezifiziert den Report von Datenqualität
- Aufbau von Informationen zur Datenqualität

Datenqualität wird als der Unterschied zwischen der zu wahrgenommenen Realität und der Wiedergabe ihrer durch die Daten definiert. Sie erhält damit einen Zweckbezug:

die Qualität von Daten aus der Sicht des Datenerzeugers kann von der aus der Sicht des Datennutzers abweichen, insbesondere wenn er die Daten für einen anderen Zweck einsetzen will.

Z. B. eine hochauflösende Biotoptypenkarte zeigt zwar auch Verkehrswege, ist aber schlecht geeignet um daraus eine Verkehrsroutenplanung abzuleiten. Der Begriff der Qualität bekommt hier einen anderen Charakter als den der Genauigkeit mit der Bestimmung von Fehlerwerten.



# ISO 19113

Data quality element Data quality sub-element	Description
Completeness	Presence or absence of features, their attributes and relationships
Commission	Excess data present in a dataset
Omission	Data absent from a dataset
Logical consistency	Degree of adherence to logical rules of data structure, attribution and relationships
Conceptual consistency	Adherence to rules of the conceptual schema
Domain consistency	Adherence of values to the value domains
Format consistency	Degree to which data is stored in accordance with the physical structure of the data set
Topological consistency	Correctness of the explicitly encoded topological characteristics of a dataset
Positional accuracy	Accuracy of the position of features
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
Temporal accuracy	Accuracy of the temporal attributes and temporal relationships of features
Accuracy of a time measurement	Correctness of the temporal references of an item (reporting of error in time measurement)
Temporal consistency	Correctness of ordered events or sequences, if reported
Temporal validity	Validity of data with respect to time
Thematic accuracy	Accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships
Classification correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data set)
Non-quantitative attribute correctness	Correctness of non-quantitative attributes
Quantitative attribute accuracy	Accuracy of quantitative attributes

# ISO 19114

**The purpose of the EN ISO 19114 is to:**

- Provide a framework of procedures for determining and evaluating quality of geographic datasets
- Establish a framework for evaluating and reporting data quality results, as part of metadata or as a data quality report

**The standard describes a general process flow to guide the data quality evaluation process. Basically it is a 6-step procedure:**

**Step 1:** Identify an applicable data quality element, sub-element and data quality scope

**Step 2:** Identify, for each sub-element and scope, a suitable data quality measure

**Step 3:** Select and apply a data quality evaluation method

**Step 4:** Determine the data quality result

**Step 5:** Determine conformance

**Step 6:** Report on results and /or conformance

- [page 20 from http://www.eurogeographics.org/sites/default/files/Guidelines\\_ISO19100\\_Quality.pdf](http://www.eurogeographics.org/sites/default/files/Guidelines_ISO19100_Quality.pdf)

# ISO 19114

## Standard classifies data quality evaluation methods:

**Direct:** can be either internal or external. An example of internal direct quality evaluation is a logical consistency test that can be performed using a dataset by itself. External direct evaluation occurs when an external dataset or the real world is used as a reference against which the dataset is evaluated.

**Indirect:** use external knowledge such as usage, lineage and purpose.

→ **Techniques for direct evaluation** are full inspection or sampling. The standard suggests using general sampling standards such as ISO 2859 and ISO 3951. Informative annex of the standard provides information on how to select appropriate sampling strategy.

[http://www.eurogeographics.org/sites/default/files/Guidelines\\_ISO19100\\_Quality.pdf](http://www.eurogeographics.org/sites/default/files/Guidelines_ISO19100_Quality.pdf)

## 2.4 TS ISO 19138 Geographic information - Data quality measure

### 2.4.1 Current status of the technical specification

Technical specification has been published in 2006.

### 2.4.2 Description of content

The objective of the technical specification is to guide the producer in choosing the right data quality measures for data quality reporting and the user in the evaluation of the usefulness of a dataset by standardising the components and structures of data quality measures and by defining commonly used data quality measures.

It defines a set of data quality measures that can be used when reporting data quality for the sub-elements in ISO 19113. The idea is to build a register of standardized quality measures. It does not limit users from defining their own quality measures.

Each quality measure is described by a set of components (see Table 2.3). Specification includes a list of data quality basic measures that can be used to describe quality measures (see Table 2.4)

# ISO 19138 - (2)

TABLE 2.3 COMPONENTS OF QUALITY MEASURES (SUMMARIZED FROM TABLE B.1 ISO 19138)

Component	Description	Obligation M=mandatory, O=obligatory, C=conditional	Comments/Examples	Component	Description	Obligation M=mandatory, O=obligatory, C=conditional	Comments/Examples
Name	Name of the data quality measure.	M					
Alias	Other recognised name for the same data quality measure.	O					
Data quality element	The name of the data quality element to which this data quality measure applies. See Chapter 2.2 on ISO 19113	M		Parameter(s)	Auxiliary variables used by the data quality measure including name, definition and description.	C	There can be one or many parameters
Data quality sub-element	The name of the data quality sub-element to which this data quality measure applies. See Chapter 2.2 on ISO 19113	M		Data quality value type	Value type for reporting a data quality result.	M	Examples include Boolean, real, integer, ratio, percentage or measure(s) (values+ units)
Data quality basic measure	Name of data quality basic measure	C	Technical specification lists a set of data quality basic measures that can be used. The user can define their own data quality basic measure. Its is typically based on counting of erroneous items, dealing with uncertainty or general statistical measures	Data quality value structure	Structure for reporting a complex data quality result	O	Bag, Set, Sequence, Table, Matrix, Coverage
Definition	Statement of the fundamental concept of the data quality measure	M		Source reference	Citation of the source of the data quality measure.	C	If an external source exists
Description	Description of the data quality measure including method of calculation with all formulae and/or illustrations needed to establish the result of applying the measure.	C	If the definition is not sufficient to understand the data quality measure concept. Example: what is not measured, not counted, what other measures should be used to help interpret the results.	Example	Example of applying the data quality measure or the result obtained for the data quality measure.	O	
				Identifier	Integer number, uniquely identifying a data quality measure.	C	If data quality measures are administered in a register



# ISO 19138 - (3)

TABLE 2.5 SELECTED IMPORTANT QUALITY MEASURES FOR THE NMCAs

Data quality element	Data quality sub-element	Data quality measure	Data quality basic measure	Identifier	Examples
Commission	Commission	Number of excess items	Error count	2	
Commission	Omission	Number of missing items	Error count	7	
Logical consistency	Conceptual consistency	Number of items noncompliant to the rules of the conceptual schema	Error count	11	
Logical consistency	Domain consistency	Number of items not in conformance with their value domain	Error count	17	
Logical consistency	Topological consistency	Number of faulty point-curve connections	Error count	22	Two roads in a junction don't meet
Logical consistency	Topological consistency	Number of missing connection due to undershoots	Error count	24	
Logical consistency	Topological consistency	Number of missing connections due to overshoots	Error count	25	
Positional accuracy	Absolute or external accuracy	Mean value of positional uncertainties (1D, 2D and 3D)	Not applicable	29	This is applicable when a set of co-ordinates considered to be true exists. $2D: e_i = \sqrt{(x_m - x_u)^2 + (y_m - y_u)^2}$ $3D: e_i = \sqrt{(x_m - x_u)^2 + (y_m - y_u)^2 + (z_m - z_u)^2}$ $\bar{e} = \frac{1}{N} \sum_{i=1}^N e_i$
Positional accuracy	Absolute or external accuracy	Mean value of positional uncertainties excluding outliers (2D)	Not applicable	30	Same as quality measure with identifier 29 except all positional uncertainties above a defined threshold are removed from the set.
Positional accuracy	Absolute or external accuracy	Mean value of positional uncertainties excluding outliers (2D)	Not applicable	30	Same as quality measure with identifier 29 except all positional uncertainties above a defined threshold are removed from the set.

Data quality element	Data quality sub-element	Data quality measure	Data quality basic measure	Identifier	Examples
Positional accuracy	Absolute or external accuracy	Covariance matrix	Not applicable	33	
Positional accuracy	Absolute or external accuracy	RMSE	Not applicable	41	Standard deviation, where the true value is not estimated from the observations but known a priori $\sigma_z = \sqrt{\frac{1}{N} \sum_{i=1}^N (Z_{mf} - z_i)^2}$
Temporal accuracy	Temporal validity	Number of items not in conformance with their value domain	Error count	17	Buildings in a dataset should have been reviewed in 2006. By mistake some of the buildings were not reviewed. Measure for temporal validity might then be number of non-valid review dates of buildings
Thematic accuracy	Classification correctness	Number of incorrectly classified features	Error count	62	
Thematic accuracy	Classification correctness	Misclassification matrix	Not applicable	64	Matrix that indicates the number of items of class (i) classified as class (j) The diagonal elements of the misclassification matrix contain the correctly classified items, and the off diagonal elements contain the number of misclassification errors.
Thematic accuracy	non-quantitative attribute correctness	Number of incorrect attribute values	Error count	67	

# ISO 19138 - (4)

TABLE 2.4 BASIC DATA QUALITY MEASURES (SEE TABLE C.1 ISO 19138, EXAMPLES ARE MODIFIED FROM THE STANDARD)

Data quality basic measure name	Data quality basic measure definition	Examples
Error indicator	Item is in error	True (item is not correct)/False (item is correct)
Correctness indicator	Item is not in error	True (item is correct) / False (item is not correct)
Error count	Total number of items that are subject to an error of a specified type	10 (number of incorrect items)
Correct items count	Total number of items that are free of errors of a specified type	200 (number of correct items)
Error rate	Number of the erroneous items with respect to the total number of items that should have been present	Error rate can be real, percentage or ratio. Note: Total number of items that should have been present should also be reported if real or percentage is used.
Correct items rate	Number of the correct items with respect to the total number of items that should have been present	See above.

# ISO 19157– Ausschnitt Schema (1)

## http://schemas.geoviqua.org/ISO/19157/20120707\_GVQ/19157\_DataQuality.xsd

Created with Liquid XML Studio 2012 Developer Edition 10.1.2.4113 (<http://www.liquid-technologies.com>)

```
-->
<xs:schema xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gmd19157="http://www.geoviqua.org/gmd19157" xmlns:gmx="http://www.isotc211.org/2005/gmx" xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:updated19115="http://www.geoviqua.org/19115_updates" xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified" targetNamespace="http://www.geoviqua.org/gmd19157" version="3.1.0">
<!-- ISO online schema location - opengis -->
<xs:import schemaLocation="http://schemas.opengis.net/iso/19139/20070417/gco/gco.xsd" namespace="http://www.isotc211.org/2005/gco"/>
<xs:import schemaLocation="http://schemas.opengis.net/iso/19139/20070417/gmx/gmx.xsd" namespace="http://www.isotc211.org/2005/gmx"/>
<xs:import schemaLocation="http://schemas.opengis.net/iso/19139/20070417/gmd/gmd.xsd" namespace="http://www.isotc211.org/2005/gmd"/>
<xs:import schemaLocation="../19139/20120707_GVQ/19115_updates.xsd" namespace="http://www.geoviqua.org/19115_updates"/>
<xs:complexType name="LI_ProcessStep_Type">
<xs:complexContent>
<xs:extension base="gco:AbstractObject_Type">
<xs:sequence>
<xs:element name="description" type="gco:CharacterString_PropertyType"/>
<xs:element name="rationale" type="gco:CharacterString_PropertyType" minOccurs="0"/>
<xs:element name="dateTime" type="gco:DateTime_PropertyType" minOccurs="0"/>
<xs:element name="processor" type="gmd:CI_ResponsibleParty_PropertyType" minOccurs="0" maxOccurs="unbounded"/>
<xs:element name="source" type="gmd19157:LI_Source_PropertyType" minOccurs="0" maxOccurs="unbounded"/>
</xs:sequence>
</xs:extension>
</xs:complexContent>
</xs:complexType>
<!--
.....
-->
<xs:element name="LI_ProcessStep" type="gmd19157:LI_ProcessStep_Type"/>
<!--
.....
-->
<xs:complexType name="LI_ProcessStep_PropertyType">
<xs:sequence minOccurs="0">
<xs:element ref="gmd19157:LI_ProcessStep"/>
</xs:sequence>
<xs:attributeGroup ref="gco:ObjectReference"/>
<xs:attribute ref="gco:nilReason"/>
</xs:complexType>
<!--
```

# ISO 19157– Ausschnitt Schema (1)

```
.....  
.....  
-->  
<xs:element name="LI_Lineage" type="gmd19157:LI_Lineage_Type"/>  
<!--  
.....  
-->  
<xs:complexType name="LI_Lineage_PropertyType">  
<xs:sequence minOccurs="0">  
<xs:element ref="gmd19157:LI_Lineage"/>  
</xs:sequence>  
<xs:attributeGroup ref="gco:ObjectReference"/>  
<xs:attribute ref="gco:nilReason"/>  
</xs:complexType>  
<!--  
=====
```

# Quality Completeness

## Completeness:

- A data set is complete with regards to aspects such as minimum area employed in polygon construction, gaps in either the data element set or attribute values, etc.
- Completeness also refers to the aspects of the data set that characterize it as a whole and not as a specific or individual element.
- Completeness can be defined in terms of commission or omission of error

## Feature completeness

defined over space and time



### Data completeness

- check on data quality
- measurable error or offset observed between the database and the specification
- if the database contains all the objects with their specifications, the data set is considered complete

### Model completeness

- agreement between the database and the abstract universe required for a specific database application

## Attribute completeness

is the degree to which all relevant attributes of a feature have been encoded



# Errors in GIS

## How they affect Data Quality

Errors encountered in **primary methods** of data collection may be classified as:

- Personal errors
- Instrumental errors
- Environmental errors

Errors encountered in **secondary methods** of data collection may be classified as:

- Errors in plotting control
- Compilation error
- Error in drawing
- Error in map generalization
- Error in map reproduction
- Deformation of the material
- Error introduced due to the use of the wrong scale
- Uncertainty in the definition of a feature
- Error due to feature exaggeration
- Error in digitizing or scanning

Additionally, all the errors encountered in primary

# Factors Affecting Data Quality

- 1) Currency
  - a) Are data up to date?
  - b) Time series
- 2) Completeness
  - a) Feature or Entity completeness
    - i) Data completeness
    - ii) Model completeness
  - b) Attribute completeness
    - i) Value completeness
- 3) Consistency
  - a) Map scale
  - b) Standard descriptions
  - c) Relevance
- 4) Accessibility
  - a) Format
  - b) Copyright
  - c) Cost
- 5) Accuracy and precision
  - a) Lineage – when collected, by whom, how?
  - b) Density of observations
  - c) Positional accuracy
  - d) Attribute accuracy – qualitative and quantitative
  - e) Temporal accuracy
- 6) Sources of errors in data
  - a) Data entry or output faults
  - b) Choice of original data model
  - c) Natural variation and uncertainty in boundary location and topology - temporal error
  - d) Observer bias
  - e) Processing
    - i) Numerical errors in the computer
    - ii) Limitations of computer representations of numbers
- 7) Sources of errors in derived data and in the results of modelling and analysis
  - a) Problems associated with map overlay
  - b) Classification and generalization problems
  - c) Choice of analysis model
  - d) Misuse of logic
  - e) Error propagation
  - f) Method used for interpolation
  - g) Lack of consistency in different analysis of the same data.

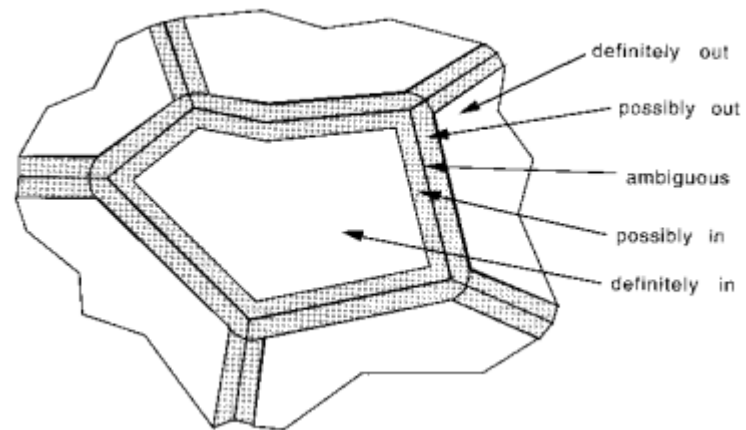
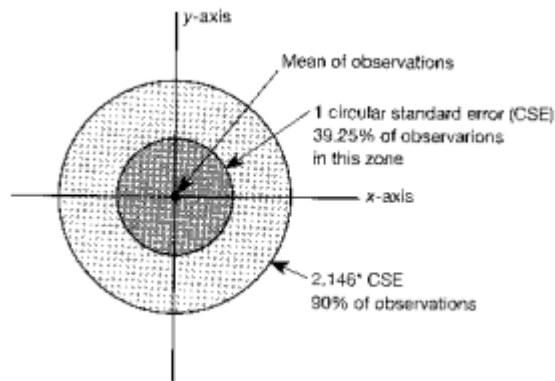
# Modelling Data Error

## Attribute Error

- can be modelled using conventional statistical techniques

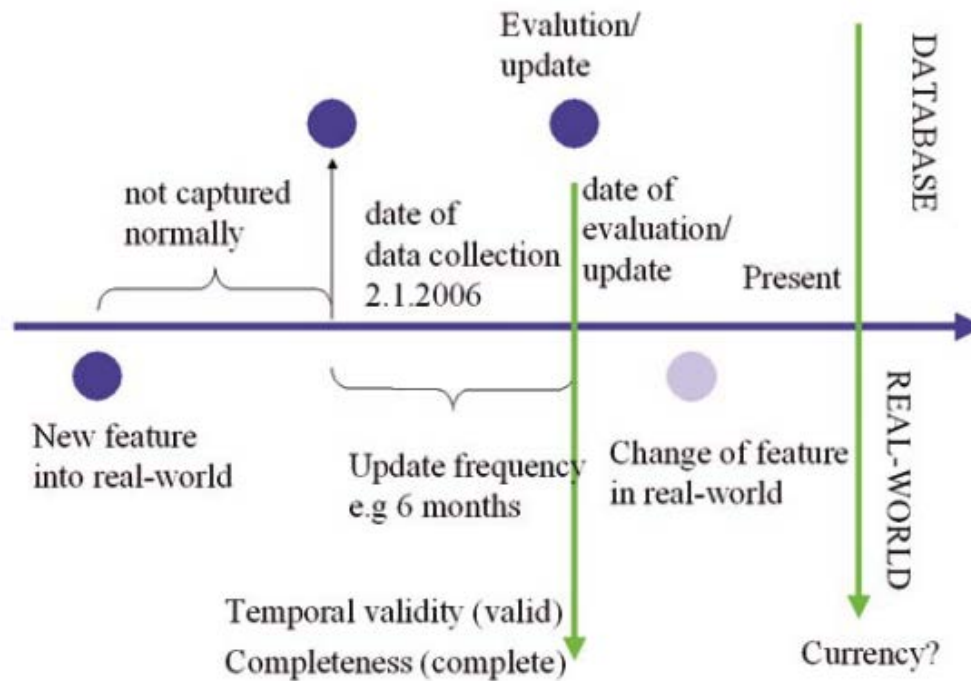
## Positional Error

- confidence bands around locational features



<http://www.nuim.ie/staff/dpringle/gis/gis11.pdf>

# Fortschreibung / Updatequalität



# OGC (Open Geospatial Consortium)

---

**WMS** is the draft Standard ISO 19128

**GML** is the draft Standard ISO 19136

**WFS** is the draft Standard ISO 19142

[http://liris.cnrs.fr/~sservign/12\\_Servigne-EN-vSS.pdf](http://liris.cnrs.fr/~sservign/12_Servigne-EN-vSS.pdf)

Jörn Kohlus (NPA) – Ulrike Kleeberg (HZG)