

# **A QUALITY MODEL FOR QUALITY ASSURANCE OF ROAD INFORMATION**

**Thomas Wiltschko & Thilo Kaufmann**

University of Stuttgart  
Institute for Applications of Geodesy to Engineering  
Prof. Dr.-Ing. Wolfgang Möhlenbrink

Geschwister-Scholl-Str. 24D – D-70174 Stuttgart  
Phone: +49 711 121-4048 – Fax: +49 711 121-4044  
E-Mail: thomas.wiltschko@iagb.uni-stuttgart.de

## **SUMMARY**

For future ITS-applications and many other areas a pan-European road data infrastructure of high quality is necessary. The EU-project EuroRoadS (Pan-European Road Data Solution) will lay the groundwork for the creation of such an infrastructure by national mapping agencies, road administrations, private enterprises and the research sector. The specific challenge regarding quality is to prepare a common model describing the quality assured data flow from acquisition and updating of the digital road map up to final application. This paper introduces a quality model of digital road maps for description of geographic data quality within acquisition and processing procedures. The quality model consists of a set of inherent quality characteristics and appropriate parameters, and is based on existing standards, like ISO 19100-series of ISO/TC211 and ISO 9000.

## **INTRODUCTION**

The actual information chain for European road data can be characterised by different national road databases from different data providers and content providers. Typically mapping agencies have developed road databases from the cartographic point of view, and road administrations have developed databases from the point of view of road maintenance. Currently no close cooperation is established between the mapping agencies, the road administrations and other parties. The data providers and content providers work at different processing steps. For the usage of national databases of information providers and service providers for European-wide applications, a common interface is necessary. At this position in the information chain the EU-project EuroRoadS will establish a specification of a road network information model, core European road data, road network exchange model, and an exchange format with the aim to create a pan-European infrastructure of road information. The specific challenge regarding quality assurance is to provide a quality concept which is usable within the entire information chain of acquisition and processing of road data (cf. figure 1). Regarding practice the applicability of the quality concept has to be ensured by all actors (data provider, content provider, information provider, and service provider). In addition, the quality concept has to be conform to the relevant standards, like ISO 19100-series.

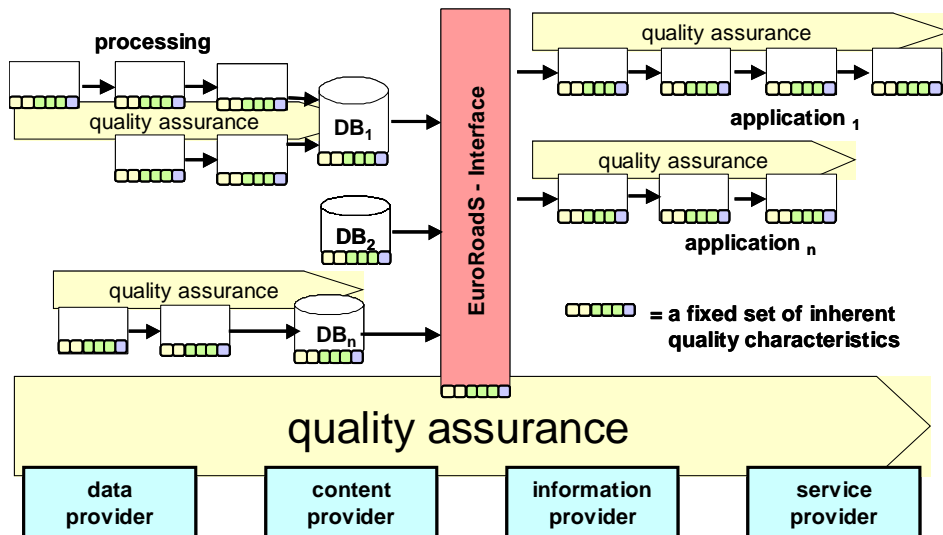


Figure 1: Information chain with a uniform quality description by a fixed set of quality characteristics as an essential precondition of quality assurance

The basic standard for quality definition is ISO 9000. Therein quality is defined as “*degree to which a set of inherent characteristics fulfils requirements*“ [1]. In this context inherent means existing in something. The note of definition emphasizes the character of a permanent characteristic. The quality characteristic in ISO 9000 is equivalent to quality element in ISO 19113 and means “*inherent characteristic of a product, process, s or system related to a requirement*”. To carry out the quality assurance within the entire information chain, EuroRoadS needs in addition to inherent quality characteristics a fixed set of quality characteristics. All actors within the information chain have to deal with identical quality characteristics in order to make the quality descriptions comparable and interoperable.

## INVESTIGATIONS OF QUALITY CONCEPTS FOR GEOGRAPHIC DATA

By the investigation of existing quality concepts in standards for geoinformation quality (ISO 19100-series, FIPS 173, DIN V ENV 12656) and in geoinformation models (GDF, ATKIS, and others) should be identified whether these quality concepts fulfil the requirements of a quality model to carry out the quality assurance and management within geoinformation processes. Tab. 1 gives an overview of specified quality elements in the various references.

The postulated requirements for intended purpose are as follows:

- One quality phenomenon must be described non-ambiguously by one quality characteristic. This is a precondition for an objective and clear quality description.
- The quality has to be described by a fixed set of quality characteristics independent of data types (geometric, thematic, temporal). The model should be usable similarly for a geographic dataset, dataset series, subsets, features, feature classes and attributes.
- Existing quality concepts and standards are to be considered.
- It should be applicable similarly for all actors.

Table 1: Quality elements in investigated quality concepts subdivided into elements with direct and non-direct quality statement

Quality elements, of which direct statement can be made:	ISO 19113	DIN V ENV 12656	FIPS 173	GDF	ATKIS
Completeness	X	X	X	X	X
Logical consistency	X	X	X		X
Positional accuracy	X	X	X	X	X
Temporal accuracy	X	X			X
Thematic/semantic/attribute accuracy	X	X	X	X	X
Resolution				X	
Precision				X	
Correctness				X	
Up-to-dateness				X	
Quality elements, of which non-direct statement can be made	ISO 19113	DIN V ENV 12656	FIPS 173	GDF	ATKIS
Lineage	X	X	X		
Homogeneity		X			
Quality documentation		X			
Purpose	X				X
Usage	X				X
History					X

For a better understanding, the investigated standards will be described: The ISO 19100-series were prepared by the Technical Committee ISO/TC211, Geographic information / Geomatics [2]. The Federal Information Processing Standards (FIPS) are issued by the National Institute of Standards and Technology (NIST) in the USA [3]. Furthermore, the European Prestandard ENV 12656 was approved by the European Committee for Standardization (CEN) as a prospective standard for provisional application [4]. The basis for digital road data in vehicle navigation systems as well as a basis for the development of ITS is the Geographic Data File (GDF) [5]. Finally, ATKIS (“Authoritative Topographic-Cartographic Information System”) is the standard for the topographic data basis in Germany [6].

The result of the investigation of these standards is to assert that the various standards and concepts show a similar structure and content as ISO 19113. In this case, the used quality elements can be divided into two types: elements of which a direct statement can be made (e.g. completeness) and elements of which only conclusions to quality can be drawn, but no direct statement is given (e.g. lineage). According to the establishing of a quality model for quality assurance in geoinformation processes, the second group can be ignored. Regarding the quality elements with direct statement, the various standards contain on the whole the same quality elements: completeness, consistency, and different kinds of accuracy.

The investigation of the quality concept in ISO 19113 has shown that the quality model cannot be used directly for the purpose of application within EuroRoadS:

- First, there is no quality element describing the phenomenon of availability of information. But especially for a service provider, availability of information is a quality characteristic of high priority.
- Second, the accuracy elements of ISO 19113 describe accuracy dependent of the data type. This means, data of various data types will be described by various quality elements. Therefore the requirement of a fixed set of characteristics is not fulfilled.
- Finally, both different phenomena of wrong digitising and limited accuracy of a digitising process will be described by one and the same quality element: *positional accuracy*. This means that the quality description using the quality elements of the ISO 19113 is ambiguous.

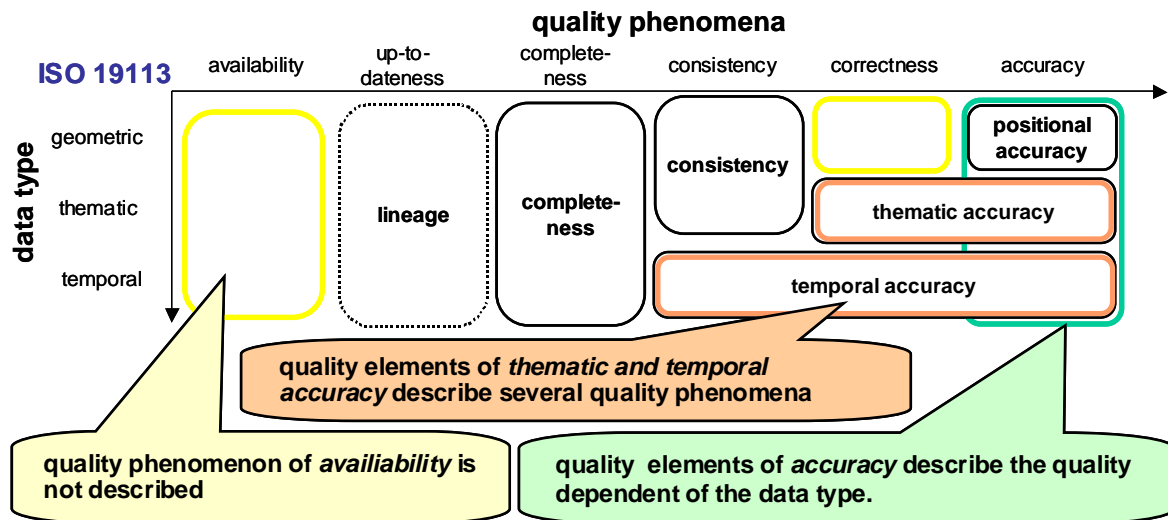


Figure 2: Identified lacks of ISO 19113

Regarding the identified lacks of ISO 19113 (cf. figure 2) the extended scope of EuroRoadS quality model should be taken into consideration. ISO 19113 contains principles for describing quality of geographic data and concepts for handling quality information for geographic data. The standard is applicable to data producers providing quality information to describe and assess how well a dataset meets the product specification. Data users can use the standard to determine whether specific geographic data have sufficient quality for their application or not [2]. The EuroRoadS quality model is applicable for all actors within the entire information chain, also for information and service providers, who are not part of the geoinformation community. In contrast to ISO 19113, the EuroRoadS quality model can carry out the quality description within geoinformation formation processes. A quality description of datasets etc. is possible as well.

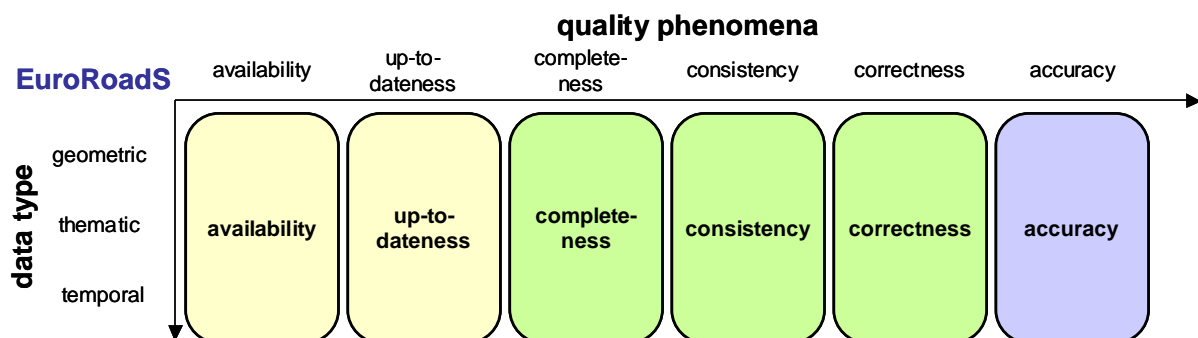


Figure 3: Quality characteristics in the EuroRoadS quality model

The defined quality model of EuroRoadS with its fixed set of inherent quality characteristics, illustrated in figure 3, represents the quality framework for the quality description of geoinformation.

## STRUCTURE AND COMPOSITION OF THE EURORoadS QUALITY MODEL

The quality model should be usable for the quality description of a geographic dataset as well as the quality description of geographic information within processes [7]. Thus an extended scope compared to ISO 19113 is given and adaptation is recommended. The use of the fixed set of inherent characteristics ensures an objective, complete, and unambiguous description of the geoinformation quality. Therefore each quality characteristic has to deal with one definite attributable phenomenon (cf. figure 4).

The EuroRoadS quality model contains six quality characteristics: availability, up-to-dateness, completeness, consistency, correctness, and accuracy (cf. figure 3). Availability and up-to-dateness are dependability characteristics describing time-related aspects of data quality. The integrity characteristics completeness, consistency, and correctness describe the fitness of use of data. Finally the accuracy: Considering that geographic data result from measurement or interpretation, or can be derived from such data and furthermore considering the fact that a measurement or interpretation can be given only with a certain accuracy or resolution, characteristics referring to this problem are essential for an information quality model.

Concretion of quality is carried out by quality parameters. A fixed set of quality parameters is not given. The use of parameters is dependent on data type and the applied method of data quality evaluation. For example *rate of omission* can be used as a parameter for the quality characteristic *completeness*. Concretion of accuracy can be effected by both parameters *standard deviation in position* and *standard deviation in height*. One important feature of the EuroRoadS quality model is that it contains all quality sub-elements of ISO 19113 as quality parameters. This means, a data provider can transfer in an easy way his present quality description according to ISO 19113 to EuroRoadS and reverse.

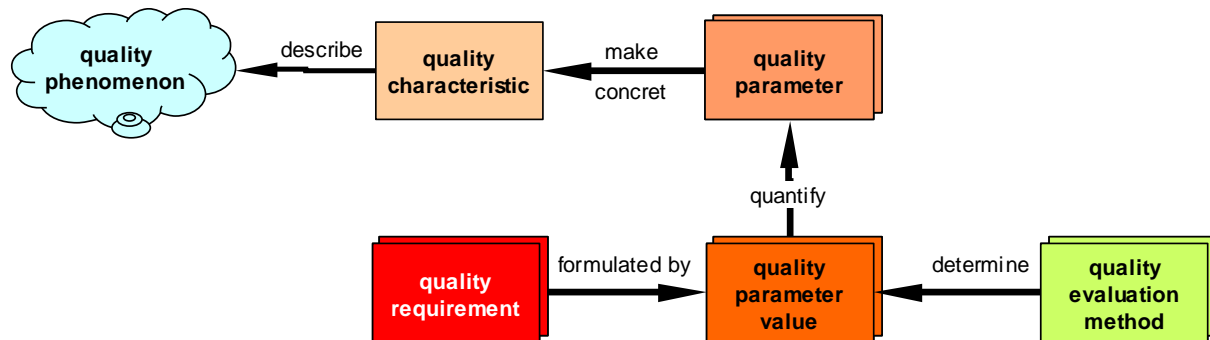


Figure 4: Structure of EuroRoadS quality model

The quantization is achieved by the parameter values which are determined by suitable quality evaluation methods. Specifying threshold values of the quality parameters by user, supplier, or another interested party carry out the formulation of quality requirements. So the following can be arranged: rate of omission less 99 %, standard deviation in position less than 5 meters and in height less than 10 meters. The first one is directed to completeness, whereas the others are directed to the accuracy of the information.

Table 2 contains the definition of the six quality characteristics and a list of possible quality parameters. The quality parameters which are also contained in ISO 19113 as quality sub-elements are labelled with (\*).

Table 2: Definitions of quality characteristics and quality parameters for the EuroRoadS quality model

groups of quality characteristics	quality characteristics	definition	possible quality parameter
dependability	availability	degree to which geographic data are available at a certain place and at a defined time	failure rate
			<i>user-defined</i>
	up-to-dateness	degree of adherence of geographic data to the time changing universe of discourse	last update
			rate of change
		temporal lapse	
		<i>user-defined</i>	
integrity	completeness	degree of adherence of the entirety of geographic data (features, their attributes, and relationships) to the entirety of the universe of discourse	omission (*)
			commission (*)
			<i>user-defined</i>
	correctness	degree of adherence of existence of geographic data (feature(s), attributes, functions, relationships) to corresponding elements of the universe of discourse, up-to-dateness being presumed	geometric correctness
			topological correctness
			thematic correctness
			<i>user-defined</i>
	consistency	degree of adherence of geographic data (data structure, their features, attributes, and relationships) to the models and schemas (conceptual model, conceptual schema, application schema, and data model)	geometric consistency
			topological consistency (*)
thematic consistency			
		<i>user-defined</i>	
accuracy	accuracy	degree of adherence of geographic data to the most plausible resp. true value.	absolute position accuracy (*)
			relative position accuracy (*)
			quantitative attribute accuracy (*)
			temporal accuracy of time measurement (*)
			<i>user-defined</i>

## Availability

The availability is only fulfilled if geographic information is available at a required place in time. In addition to the time factor, the place is also a fundamental importance. It is not sufficient that geographic data exist, the geographic data must also be available at the required place and time. Therefore the availability can be defined as “*degree to which geographic data are available at a certain place and at a defined time*”. For example, if in case of a WebGIS-application the geographic data is available in the intranet, but due to a communication error cannot be transferred to the internet to external users, availability is not given. I.e. if the system or a part of it is failed as a result of a technical failure, availability is not given.

In many cases the *failure rate* of the technical component can be used as quality parameter of the availability, because if a sensor or another technical component of the information chain fails, the information delivered by this sensor will be non-available as well. The failure rate indicates the degree of probability of a failure within an assumed period and can be measured by the equation

$$R(t) = \frac{n_{AV}(t)}{N} \quad (1)$$

whereby the quotient is the number  $n_{AV}$  of missing entities at time  $t$  and number  $N$  of entities in system.

## Up-to-dateness

The up-to-dateness describes the degree of adherence of geographic data to the time changing universe of discourse. The universe of discourse is a model of the real world which contains the pieces human selects for description. This is the view of the real or hypothetical world that includes everything of interest [8]. The up-to-dateness depends on the dynamic of reality and the frequency of update. For example, flow regulation at a junction was changed in March 2005 from traffic sign to traffic light. But the regulation of the junction was not changed in the geographic dataset till now. As a result of changes in the course of time, the information does not reproduce the reality correctly. Between up-to-dateness, completeness, correctness, and availability exist strong connections which have to be considered at the observation of the information quality within information processes.

Other examples for the influence from up-to-dateness are illustrated in figure 5. In scenario A the geometry of a road element has changed, but there is no current maintenance of the database. The consequence of this missing update is a commission. Furthermore, the calculated length of the route will be incorrect. In scenario B roads were built in reality, but there is no current maintenance of the database. The missing update will lead to an omission. Furthermore, the routing in this road network cannot be calculated, the route is unavailable.

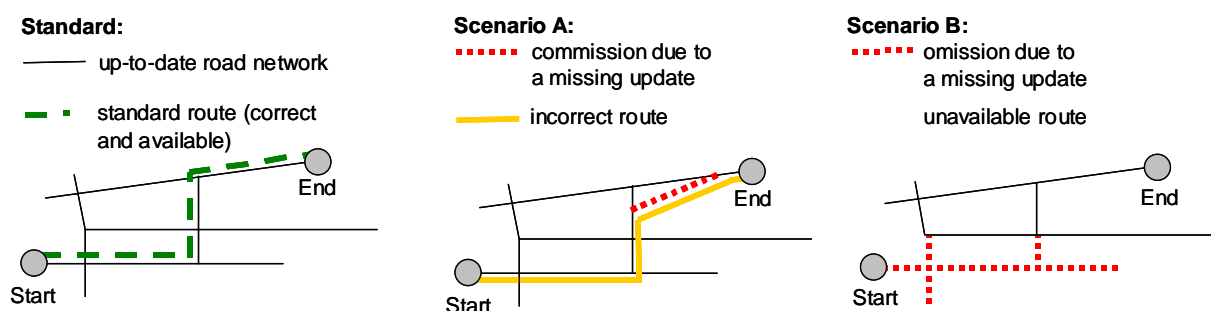


Figure 5: Examples for influence of up-to-dateness

## Completeness

The applicability resp. the fitness of use of geoinformation is described by the integrity characteristics completeness, consistency, and correctness. The completeness is fulfilled if all required features, attributes and relationships of the universe of discourse are existent in the geographic dataset. For example if the feature *Point of Interest* consists of coordinates for geo-referencing and a name according to a specification, and the name is missing the dataset

is incomplete. This phenomenon can be formulated by the quality parameter *omission*. On the other hand an excess data can be present in a dataset, which will be described by the parameter *commission*.

One possibility to carry out quantification of the integrity characteristics is the failure rate  $\lambda$  concerning completeness (CM), consistency (CN) or correctness (CR):

$$\lambda_{\text{CM}} = \frac{n_{\text{CM}}}{N}, \quad \lambda_{\text{CN}} = \frac{n_{\text{CN}}}{N}, \quad \lambda_{\text{CR}} = \frac{n_{\text{CR}}}{N} \quad (2)$$

whereby  $N$  is the number of entities in the universe of discourse and  $n_{\text{QC}}$  the number of faulty entities according to treated quality characteristic  $\text{QC}$ . In terms of a positive logical, the quantification can be achieved by the rate of fulfilment:

$$r_{\text{CM}} = 1 - \lambda_{\text{CM}} = 1 - \frac{n_{\text{CM}}}{N}, \quad r_{\text{CN}} = 1 - \lambda_{\text{CN}} = 1 - \frac{n_{\text{CN}}}{N}, \quad r_{\text{CR}} = 1 - \lambda_{\text{CR}} = 1 - \frac{n_{\text{CR}}}{N} \quad (3)$$

## Consistency

Consistency is a precondition for a contradiction-free database and can only be described by a Boolean value for a single information (i.e. a feature, attribute, or rule are consistent or not). For a whole dataset etc. the consistency can be carried out by a probability value. The same will be done with the other quality characteristics except accuracy. A verification of consistency can be performed according to the conceptual and data model. Compared to all other quality characteristics no reference data are necessary. On the level of quality parameter, the consideration of consistency can be divided into geometric, topological and thematic aspects as recommended by the ISO 19113.

For a quality observation within a data providing process, in many cases a transition of inconsistency to failure of availability, completeness, and correctness occur and have to be taken into consideration.

## Correctness

The correctness means the extent of conformity of geographic data in relation to the universe of discourse and considers gross and systematic errors, like they mostly arise during the information formation phase. For metric and geometric data the determination of threshold, i.e. whether an incorrectness is given, depends on the existing accuracy of the measuring method or system.

Figure 6 illustrates for (I) and (II) an incorrect road geometry as a result of inaccurate digitising. Road (III) does also not correspond exactly with the road geometry in reality. Because the deviation is smaller than the accuracy interval, the collection of the road geometry is correct. In this case the deviation follows from unavoidable randomise errors.

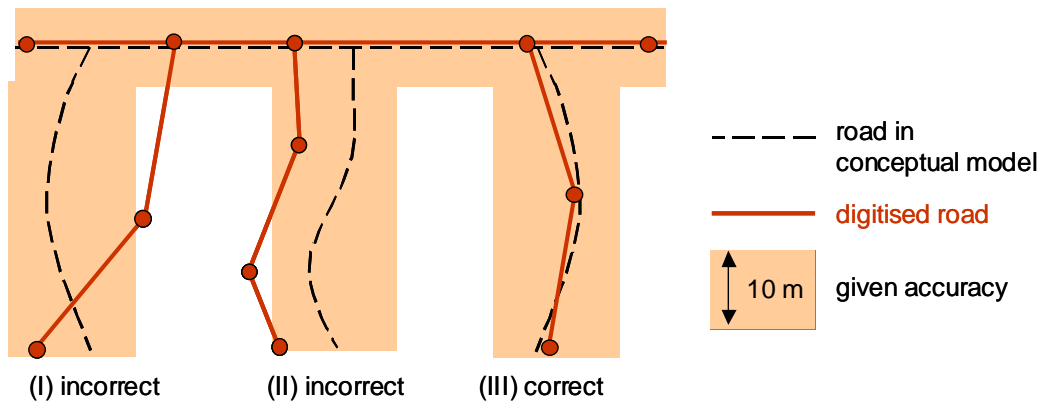


Figure 6: Incorrectness because of (I) gross and (II) systematic errors at digitising of road geometry

## Accuracy

In the same way as for the consistency one has to separate among different data types ((geo-)metric, temporal, thematic). This has to be carried through by means of different quality parameters. For all measurable and continuously represented values the accuracy can be declared by standard deviation, variance, or confidence ellipse etc. For example the accuracy of vehicle position measured by GPS can be determined with 12 meters horizontal and 25 meters for the height.

For discrete resp. discontinuous data (e.g. road class type) it is not possible to determine the accuracy by a single value. The level of discretisation depends on the degree of generalisation during the specification of the universe of discourse, and of the conceptual and the data model. In most instances the fulfilment of this kind of accuracy has to be clarified in connection with the question whether the dataset or database contains the required information and information density for the intended purpose. For example an application needs the geometry of each lane and the digital map contains only the roads, then the data source is not as detailed as necessary and consequently not suitable.

## QUALITY ASSURANCE OF ROAD INFORMATION

### Safe information

An information processing system is a system that uses different kinds of data and information to generate information for a concrete purpose of use by a defined process. If a faulty information is provided because of an imperfect input information or information processing, a faulty system behaviour will be the consequence. To avoid a hazard for humans or things within the system environment, a safe operation of the system has to be ensured. For an information processing system, the requirement of safe operation confers on the information provided by the system.

Due to the fact that no absolutely safety can be achieved in reality, the risk based on the system has to be limited on an accepted level. In this regard, safety information are necessary which exhibit a quality satisfying the required safety level. To reach the required information

quality, quality assurance measures with regard to error avoiding and error control has to be defined and implemented within the information process. The quality concept, illustrated in the following cause, is a helpful method to analyse the information flow and to design and develop quality assurance measures.

## Quality concept for information

The quality concept for information, developed and applied in EuroRoadS, consists of two parts (cf. figure 7). One part is the quality, illustrated further up, which contains the set of quality characteristics and the quality parameter, and represents the framework for the quality description of information. The other part is the analysing procedure, which includes the information flow chart and a computing procedure. The information flow chart is graphical representation and is used for modelling the information flow of the system or process. The computing procedure is used for the determination of the information quality within the information flow by evaluation the information flow chart [9].

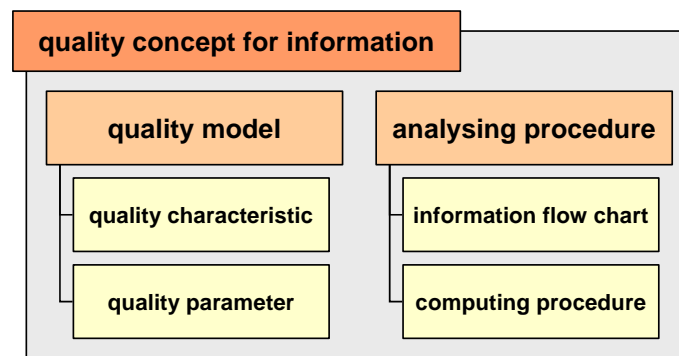


Figure 7: Structure of quality concept for information

The analysing procedure is based on the known method of reliability analysis and is derived from the fault tree and event tree analysis. Both methods are based on Boolean modelling and contain graphical and analytical parts. Compared to the reliability analysis, which only use the reliability, the analysing procedure for information has to deal with the six quality characteristics simultaneously. The probability of fulfilment the quality requirement of each quality characteristic is used as calculation quantity. The occurred dependencies, like they illustrated further up, can be modelled and taken into consideration in the computing part.

## Application of quality assurance in the PDCA-cycle

The quality assurance of road information which is described above, can be integrated into the so-called PDCA-cycle. This cycle which include the four components *Plan, Do, Check and Act* (PDCA), was originally conceived by Walter Shewhart in 1930`s, and later adopted by W. Edward Deming. The model provides in general a framework for the improvement of a process or system [10].

In the phase *Plan* objectives and processes necessary to deliver results in accordance to a specification will be established. For this purpose, the required quality of road information can be described by the fixed set of quality characteristics and formulated by variable quality parameters.

Furthermore, into the phase *Plan* the simulation of processes by using the analysing procedure can be included. I.e. additional to the cycle for real processes a second cycle for simulated processes can be useful. This includes firstly the modelling of data providing processes. On basis of the model of the data providing process, quality lacks in the process can be identified and quality assurance measures can be designed and modeled. Their impact on quality can be evaluated by the analysing procedure. In this procedure the new parameter values in consequence of the simulated assurance measures will be calculated and compared with the required parameters values. If this check will lead to the result, that the required quality is reached, the simulated assurance measure can be implemented in real processes. Otherwise, if the quality requirements are not fulfilled, additional measures have to be simulated in a second iteration. For all this phases of simulation the analysing procedures, existing of information flow chart and computing procedure, can be a helpful tool.

After the planning, including the described simulation, the phase *Do* follows. This contains an implementation of measures in a real data providing process. The following phase *Check* includes monitoring and evaluating the processes and results against the objectives and specifications. The outcome can be documented in a quality report. Therefor the evaluation methods for measuring quality parameter values can be used. Afterwards it has to be verified whether the required product quality is fulfilled. The modifications have to be formalised in a quality report using the quality framework (*Act*) (s. figure 8).

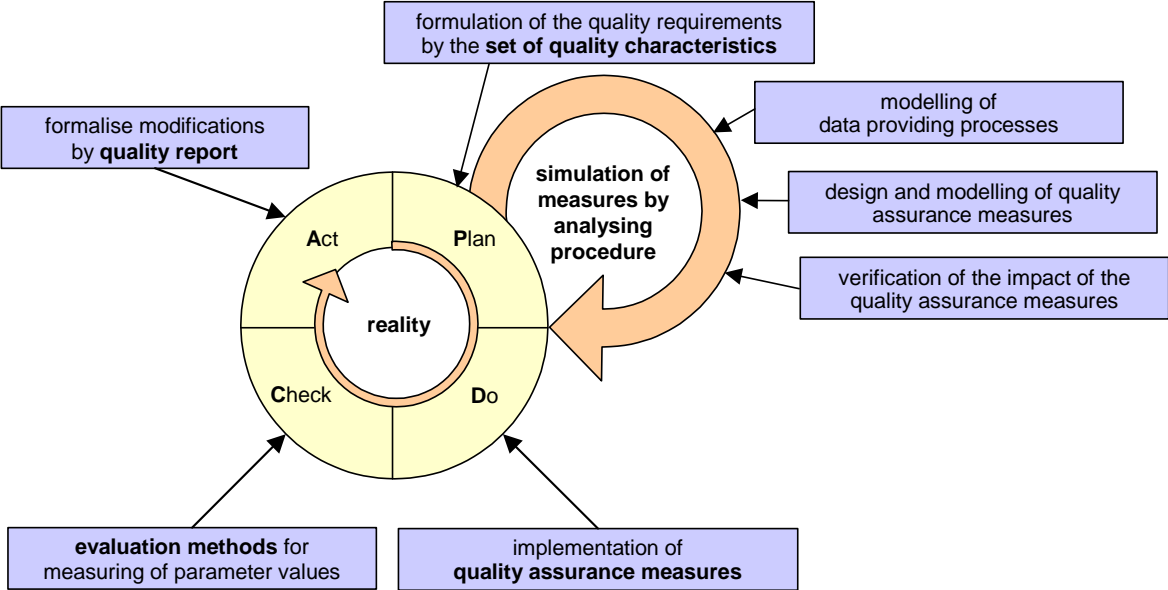


Figure 8: Integration of EuroRoadS quality concept into the extended PDCA-cycle

## CONCLUSIONS

The presented quality model with its fixed set of six inherent quality characteristics and a collection of quality parameters is part of the quality concept for geographic data (s. figure 7). The quality model is also basis for an analysing procedure allowing the quality description within geoinformation processes. The analytic part is effected by a probabilistic model based on the reliability analysis of mechanical engineering. Information flow chart will be used for modelling the information chain.

In the EuroRoadS project an analysing procedure describing and evaluating quality of geographic data within the entire information chain will be developed. The actual data quality and the provision of measures for quality assurance of geographic data will be applied in a test side evaluation.

The use of the illustrated quality concept is not restricted to geoinformation processes or geoinformation systems. The method is directly transferable on each information and information process occurring within ITS-applications [11].

## REFERENCES

- [1] CEN – European Committee for Standardization: *EN - ISO 9000: Quality management systems – Fundamentals and vocabulary*, 2000.
- [2] ISO - International Organization for Standardization: *ISO 19113: Geographic information – Quality principles*, 2002.
- [3] NIST - National Institute of Standards and Technology: *Federal Information Processing Standards Publication 173-1. Announcing the Standard for Spatial Data Transfer Standard (SDTS)*, 1994. ([www.itl.nist.gov/fipspubs/fip173-1.pdf](http://www.itl.nist.gov/fipspubs/fip173-1.pdf))
- [4] DIN – Deutsches Institut für Normung e.V.: *DIN V ENV 12656. Geoinformation – Datenbeschreibung – Qualität; Deutsche Fassung ENV 12656*, 1998.
- [5] CEN TC 278: *Geographic Data Files. Draft of a European Standard*, 1995. ([www.ertico.com/links/gdf/gdfdoc/gdfdoc.htm](http://www.ertico.com/links/gdf/gdfdoc/gdfdoc.htm))
- [6] ADV - Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland: *Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok)*. Version 3.0. Bonn: 2004. ([www.adv-online.de](http://www.adv-online.de))
- [7] Wiltschko, T.; Kaufmann, T.: *Quality frame for information*. EuroRoadS project deliverable D2.2. Stuttgart: 2004. ([www.euroroads.com](http://www.euroroads.com))
- [8] ISO - International Organization for Standardization: *ISO 19101 Geographic information – Reference Model*. 2002.
- [9] Wiltschko, T.; Kaufmann, T.: *Probabilistic model to describe and evaluate information quality*. EuroRoadS project deliverable D2.3. Stuttgart: 2005. ([www.euroroads.com](http://www.euroroads.com))
- [10] Deming, W.E.: *Out of the crisis*. Massachusetts Institute of Technology, Centre for Advanced Engineering Study, 1982
- [11] Wiltschko, T.: *Sichere Information durch infrastrukturgestützte Fahrerassistenzsysteme zur Steigerung der Verkehrssicherheit an Straßenknotenpunkten*. Dissertation, Fortschritt-Bericht VDI, Reihe 12, Nr. 570. Düsseldorf: VDI-Verlag, 2004.