Probabilistic model to describe and evaluate information quality

Project acronym: EDC-11145 EUROROADS/28646
Deliverable: D 2.3.
Nature: Public
Author: Thomas Wiltschko, Thilo Kaufmann, University of Stuttgart
Date: 18/02/05
Version: Final

Document control:

Version	Date	Editor	Comment
WD 0.1	10/12/04	Thomas Wiltschko, Thilo Kaufmann	First working draft sent to WP2 for comments
WD 0.2	24/01/05	Thomas Wiltschko, Thilo Kaufmann	Updated after internal review by members of WP2
FD 0.1	24/01/05	Thomas Wiltschko	Final draft sent to OS, VV and BEV for review by PMG
FD 0.2	18/02/05	Thomas Wiltschko, Thilo Kaufmann	Updated after review by PMG
F 1.0	18/02/05	Thomas Wiltschko	Final sent to PM to send to EC

Table of Content:

1 Introduction

1.1 Aims and scope

In work package 2.3 of the EU-project EuroRoadS a probabilistic model to describe and evaluate quality of geoinformation within data providing processes has to be specified. For this purpose a method is derived from the reliability analysis method of mechanical engineering. It will be used for the description and modelling of data providing processes and for the evaluation of the information quality within such processes.

1.2 Methodology and sequence

The EuroRoadS deliverable D2.3 “Probabilistic model to describe and evaluate information quality“ is structured as follows: First the general structure and composition of the EuroRoadS quality concept are explained and the quality model with its set of quality characteristics and parameters, which is part of the quality concept and is explained in detail in WILTSCHKO & KAUFMANN (2004), will be introduced (chapter 2). The analysing procedure is based on reliability analysis methods, which are explained in chapter 3. Based on these general descriptions of reliability analysis methods a specific method for describing and evaluating information quality in data providing processes is derived. This method consists of a graphical information flowchart as well as a computing procedure, based on probabilistic methods (chapter 5). In chapter 6 the basic and typical processes of data acquisition and processing will be described and modelled by use of the derived method. Finally the main results are summarised and an outlook is given (chapter 7).

2 Quality concept

The specific challenge is to prepare a common concept describing and validating the quality assured data flow from acquisition and update of geoinformation up to final applications. Geoinformation within the data providing process as well as at the end of the process are a product in term of ISO 9000.

Please note that a differentiation into the terms ”information” and ”data” is not always practicable. Therefore in the following the two terms will be used similarly.

The EuroRoadS quality concept consists of the quality model, which is described in the published deliverable D2.2, and an analysing procedure, which is described in this report.

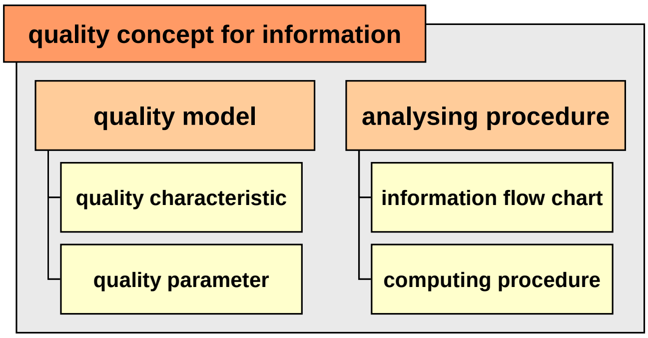


Fig. 1: Structure of EuroRoadS quality concept

2.1 Quality model

In deliverable D2.2 (WILTSCHKO & KAUFMANN 2004) a quality framework for the description of the information quality is specified. In opposite to ISO 19 113, which has scope of quality description of geographic dataset, the quality framework for EuroRoadS has the aim of describing and evaluating the quality in the entire information chain. For this reason an adaptation and extension of the ISO 19 113 was necessary.

The quality model use quality elements and quality subelements like ISO 19 113. But for clear notational separation from ISO 19 113, the quality elements are termed as quality characteristics and the quality subelements as quality parameters. The EuroRoadS quality model contains six quality characteristics:

- availability
- up-to-dateness
- completeness
- consistency
- correctness
- accuracy

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 characteristic is essential for an information quality model, because geographic data results in any way from measurement or interpretation, which can be carried out with limit precision. The six quality characteristics established the fixed set of inherent quality characteristics, which are necessary for an objective and uniform description of quality of all geographic data within the entire information chain.

Tab. 1 includes a list of possible quality parameters, which are used for a concretion of the quality (e.g. a quality quantity) as well as for formulation of quality requirements. 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*. The quality requirement can now be formulated as following: rate of omission less than 99%, standard deviation in position less than 5 meter and in height less than 10 meter.

Tab. 1: Definition of quality characteristics and quality parameters for EuroRoadS quality model for quality assurance within geoinformation processes

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
			…
	up-to-dateness	degree of adherence of geographic data to the time changing universe of discourse	last update
			rate of change
			temporal lapse
			…
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
			…
	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
			…
	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
			…
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
			…

The essential properties of the specified quality model of EuroRoadS can be summarized as following:

- One quality characteristic describes unambiguously one quality phenomenon. This property is a important precondition for a clearly quality description.
- By using a fixed set of quality characteristics independent of data type (geometric, thematic, temporal) and of processing step (data, content, information, service provider) and objective quality description is given.
- The quality model can be used in the same way for geographic datasets, data series, subsets, feature classes, features and attributes.
- The quality concept considers existing quality concepts and standards as far as possible. So an unique relationship between the quality elements and subelements of ISO 19 113

and the specified quality characteristics and parameters is given, which allows an transfer of quality description between the both concepts.

- The quality concept is applicable for all participants within in the information chain.

2.2 Analysing procedure

The aim of deliverable D2.3 is to specify a method to describe and evaluate information quality within geoinformation processes. For this analysing procedure the following requirement are postulated:

- Intuitive understandable modelling of the information flow in form of a graphical representation, which illustrates all processing steps and components with influence on information quality.
- Possibility of a modular structure of the information processes.
- Evaluation of the information flow by means of probability calculation.
- Usable for the analysis and evaluation of existing processes as well as development and design of future information processes.
- An interdisciplinary applicability of the analysing procedure so that every involved party can be participate and support the process.

The method consists of a graphical information flowchart and a computing procedure, based on probabilistic calculation (see Fig. 1)

3 Reliability analysis methods

Several methods of safety and reliability analysis exist in mechanical engineering, plant engineering, aerospace etc.. Fig. 2 shows an assortment of a part of existing methods. Furthermore other methods, which are based on petri net, neural networks, fuzzy-theory, etc., are used.

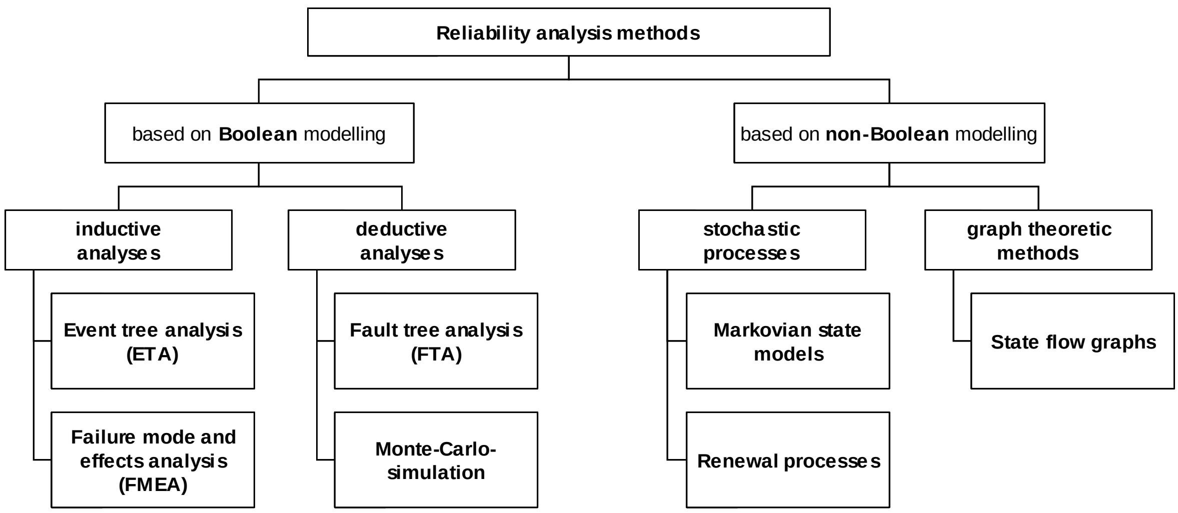


Fig. 2: Overview of an extract of reliability analysis methods (according to MEYNA (1982))

The methods can be divided according to different criteria:

- Methods, which are based on Boolean modelling and methods, which are not based on Boolean modelling.
- The Boolean methods can be divided in deductive and inductive methods. Deductive methods start at an undesirable output event (= failure of the system) and deal with input events and their combinations (= failure of system components), which can cause it. Inductive methods deal with a failure of a system component and its effects on the system.
- Kind of evaluation: graphical, analytical or textual. Fault tree analysis and event tree analysis contain a graphical and analytical part. Monte-Carlo-Simulation is analytical method for evaluation of complex systems, which cannot be evaluated by other methods. The failure mode and effects analysis is a systematic approach that identifies potential failure modes in a system. Therefore the failure type, failure probability, and the effects on the system are collected in a table.
- Application range: project planning, implementation, and application of the system.

For describing and evaluating the information chain, we need a method, which affords graphical modelling of the information flow as well as analytical evaluation of the information quality within the information flow. The method should be easy to apply and applicable to complex systems, too. The applicability should cover all application ranges.

According to the denoted requirements the Boolean models are most suitable. The probabilistic model for describing and evaluation the information quality within information processes will base on the fault tree analysis and event tree analysis, which both contain a graphical description and an analytical evaluation. For the calculative analysis the Monte-Carlo-simulation is also suitable. Within the scope of this report, the applicability of the Monte-Carlo-simulation will be disregarded.

4 Mathematical basis of Boolean model

The Boolean algebra is the mathematical basis of the Boolean model. The condition of a component is indicated by binary condition variable. In case of positive logic it is

$$X_i = \begin{cases} 1 & \text{if system is operating} \\ 0 & \text{if system failed} \end{cases} \tag{4-1}$$

Boolean operations are used to link the components. Following operations are separated:

- Conjunction (AND, $\wedge$)
- Disjunction (OR, $\vee$)
- Complement (NOT, $\bar{\ }$).

The assessment of the operations do not effected by algebraic arithmetic operations, but according to the truth table illustrated in Tab. 2. In dependence of the conditions of the system components and its logical linkage, the condition of the system can be concluded. For that purpose, the structure function is used, which state system’s operational efficiency in dependence of the conditions of the system components as a two-valued function of the condition vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. Therefore

$$\varphi(X_1, X_2, \ldots, X_n) = \begin{cases} 1 & \text{if system is operating} \\ 0 & \text{if system failed} \end{cases} \qquad (4\text{-}2)$$

is valid.

Tab. 2: Truth table

Variable		Conjunction	Disjunction	Complement
X_1	X_2	$X_1 \vee X_2$	$X_1 \wedge X_2$	$\overline{X}_2$
0	0	0	0	1
0	1	1	0	0
1	0	1	0	
1	1	1	1	

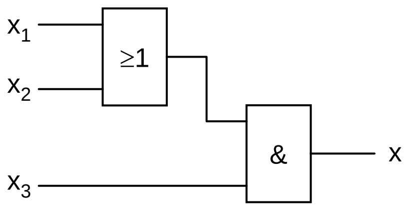
In practice, the linkage of the condition variables is replaced by suitable arithmetic operations. By usage of the real algebraic notation the following algorithm count for the operations:

- AND: $X_1 \wedge X_2 = x_1 \cdot x_2 = x_1 x_2$ (4-3)
- OR: $X_1 \vee X_2 = x_1 + x_2 - x_1 \cdot x_2 = x_1 + x_2 - x_1 x_2$ (4-4)
- NOT: $\overline{X}_1 = 1 - x_1$ (4-5)

Fig. 3 shows an example for the description of a logical connection in the form of a logic diagram. The respective structure function is:

$$\begin{aligned} \varphi(\boldsymbol{X}) &= (X_1 \vee X_2) \wedge X_3 \\ &= (x_1 + x_2 - x_1 x_2) x_3 = x_1 x_3 + x_2 x_3 - x_1 x_2 x_3 \end{aligned} \qquad (4\text{-}6)$$

According to (4-6) the system is operating if component C_2 or C_2 is operating and the component C_1 is operating at the same time.



condition table:

x_1	0	1	0	0	1	1	0	1
x_2	0	0	1	0	1	0	1	1
x_3	0	0	0	1	0	1	1	1
x	0	0	0	0	0	1	1	1

Fig. 3: Logical diagram with its system conditions

The switch to the probability realised by replacing the Boolean quantity by a discrete two-valued random quantity with the probability of operational efficiency $P\{X_i = 1\}$ and the probability of failure $P\{X_i = 0\}$ of the component C_i. The expected value $E\{X_i\}$ of the operational ability of the component C_i is calculated as follows:

$$E\{X_i\} = 0 \cdot P\{X_i = 0\} + 1 \cdot P\{X_i = 1\} \qquad (4\text{-}7)$$

If we substitute $X_i = p(x_i)$ and $\varphi(\boldsymbol{X}) = p(\boldsymbol{x})$ in (4-6), we can calculate the probability of an operating system in dependence of the probability of the operative ability of the components.

If we suppose for the both components C_1 and C_2 a probability of operative ability of $p(x_1) = p(x_2) = 0{,}99$ and for the component C_3 of $p(x_3) = 0{,}999$ of the example above, the probability of an operative system is calculated to:

$$p(\boldsymbol{x}) = (p(x_1) + p(x_2) - p(x_1) \cdot p(x_2)) \cdot p(x_3) \approx 0{,}9989 \quad (4\text{-}8)$$

Consequently the Boolean model allows the calculation of the probability of an operative system by given probabilities of the operative ability of the system components and accomplishments about the kind of linkage of the components. In the following chapter the methodology will be transferred to the task to evaluate the information quality within information processes.

5 Method to describe and evaluate information quality

5.1 Introduction

The aim is to provide a method, which describes the quality of geoinformation within the entire information chain. Therefore an analysis will be developed, that contains:

- a graphical part for the representation of the information flow (see chapter 5.2) and
- a computational part for evaluating the information quality within the information flow (see chapter 5.4)

The method is derived from the methods applied for reliability analysis in mechanical engineering.

5.1.1 Annotation

In the following several essential properties are presented, which describe the difference to the methods of reliability analysis methods of mechanical engineering (see chapter 3):

- The calculation quantity is the probability of fulfilling the quality requirements for the quality characteristics.
- In opposite to the reliability analysis, that deals with the reliability only, here several quality characteristics has to be investigated simultaneously.
- As shown in chapter 5.3 dependencies between the quality characteristics exist, if we consider information in flow. These dependencies have to be taken into account.

5.1.2 Definitions and Notations

For the calculation expressions the following notations are used. The notation of the quality characteristics is effected by:

AV .. availability	CN ... consistency	CR ... correctness
CM .. completeness	UP ... up-to-dateness	AC ... accuracy

The six quality characteristics in the given sequence defines the ordered set of quality characteristics:

$$\boldsymbol{QC} = \left\{QC(q)\,\middle|\,1 \le q \le 6\right\} = \left\{\mathrm{AV}, \mathrm{CM}, \mathrm{CN}, \mathrm{UP}, \mathrm{CR}, \mathrm{AC}\right\} \qquad (5\text{-}1)$$

By the quality values, which describe the probability of fulfilment the requirement of a quality characteristic, the following symbols are separated:

$I^{QC(q)} \equiv p(I, QC(q))$	Probability of fulfilment of the requirement of the treated quality characteristic $QM(q)$ of the input information I (for q = 1..6)
$O^{QC(q)} \equiv p(O, QC(q))$	Probability of fulfilment of the requirement of the treated quality characteristic $QM(q)$ of the output information O (for q = 1-6)

The quality of information is described by a quality tuple

$$\boldsymbol{I} = \left(I^{AV}, I^{CM}, I^{CN}, I^{UP}, I^{CR}, I^{AC}\right) \text{ resp. } \boldsymbol{O} = \left(O^{AV}, O^{CM}, O^{CN}, O^{UP}, O^{CR}, O^{AC}\right) \qquad (5\text{-}2)$$

The prerequisite for the application of the system function is the stochastic independence of the input information. In this context independence means, that the fulfilment (or non-fulfilment) of the quality requirement of any input information does not influence the quality of another input information. If this is not the case, the following expressions are not allowed to be used.

Within the information flow, there exist different feasibilities of dividing the information flow in different paths. By these paths disjunctive system statuses are modelled. For a distinction a numerical value describing the probability of the system status will be illustrated in brackets and cursive.

5.2 Graphical part for the representation of the information flow

5.2.1 General remarks for graphical symbols

The following two objects are most important for the graphical part:

- information (e.g.: street network)
- application: an object in which input information enter and output information goes out (e.g.: matching & merging). The detail information, in which way the input information are processed to generate the output information is not necessary. For modelling the information flow the kind of logical connection is sufficient (e.g. conjunction or disjunction).

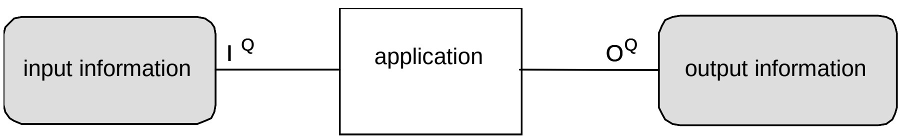


Fig. 4: Application including input information and output information as basic elements of information flowchart

The different applications are described by the graphical symbols explained in chapter 5.2.2 and illustrated using simple examples.

5.2.2 Graphical symbols

AND-linkage

The AND-linkage is the logical average for two or more input information. These linkage type is used for applications, which need all the input information for the derivation of the output information. An example for an AND-linkage is illustrated in

Fig. 5. In this example a road network and speed traffic signs will be merged and matched together. The output information can be compiled only, if both input information fulfil the quality requirements.

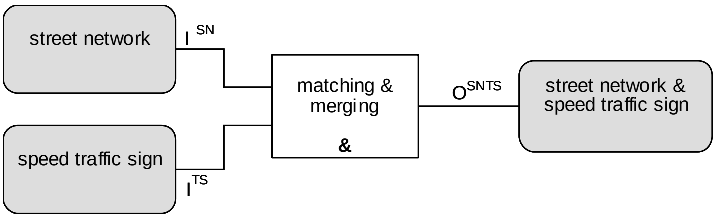


Fig. 5: Example for AND-linkage

OR-linkage

The OR-linkage is used for the modelling of redundancy. This linkage will be used, if at least one input is necessary for the generation of the output information. An example for an OR-linkage is illustrated in Fig. 6. In this example three datasets will be used as input information. At least one input information has to be available to generate the output information.

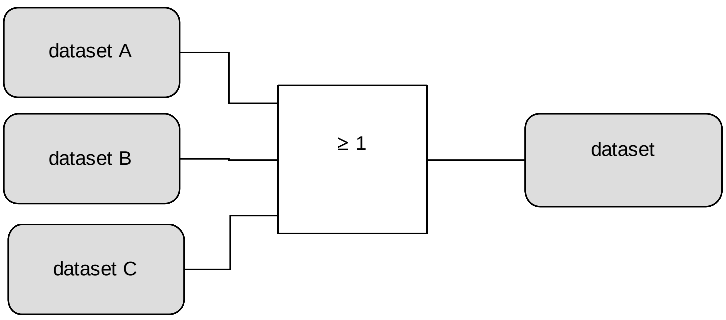


Fig. 6: Example for OR-linkage

Single branching

The single branching is an important application for modelling branches within information flow. The branching take place in dependence of a specific criteria. In the illustrated example the street network will be split into the areas inside of agglomerations (urban areas) and outside of agglomerations (interurban areas).

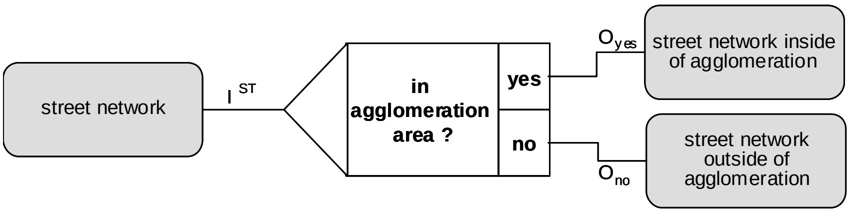


Fig. 7: Example for single branching

Single check

The single check contains a quality control within the information process. In succession of single check a split of the information flow into two paths is given. At this one path contains the checked information and the other path indicate the probability that no information is availability. Because there is no information flow in the NO-path, the flow is illustrated by an arrow. In the example in Fig. 8 a check regarding correctness (CR) is illustrated. If the dataset do not fulfil the correctness, the information will be eliminated. The result of the checked dataset is on the one side, that the correctness is increased due to the check (YES-path). On the other side the completeness (CM) is reduced due to the elimination of data, which are identified as incorrect.

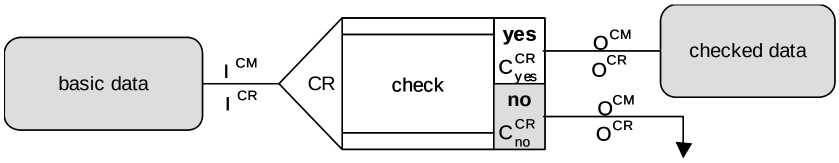


Fig. 8: Example for single check

In many cases, the data which do not fulfil the requirements will be reworked. This can be modelled by a new information processes starting at the arrow symbol.

Exclusive OR-linkage

The exclusive OR-linkage is used for the connection of disjunctive system states, who they occur as a result of a single branching or single check. In the exemplary Fig. 9 an exclusive OR-linkage of checked basic data and reworked data is illustrated.

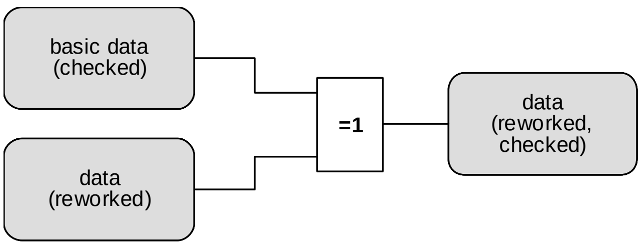


Fig. 9: Example for exclusive OR-linkage

Transition

The transition symbol is used to describe explicit the influence of one quality characteristic on another quality characteristic. The symbol occur in combination with an application. It is also possible to use the symbol for modelling the influence of a specific quality characteristic. In the exemplary Fig. 10 the influence of dependency of time on the usage of digital map is illustrated. I.e. the up-to-dateness (UP) of the digital map changes. The detailed relations among the quality characteristics, like they occur within the information flow, will be explained in chapter 5.3.

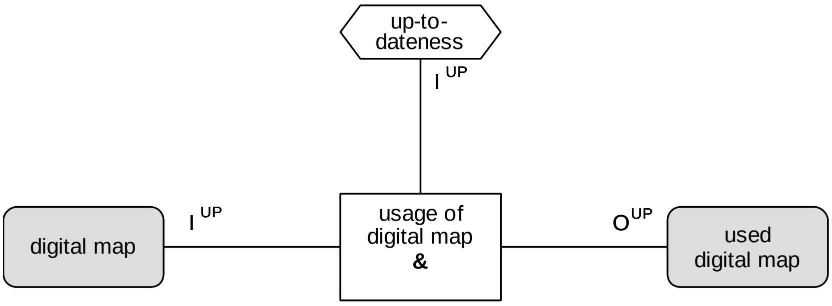
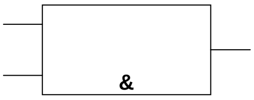
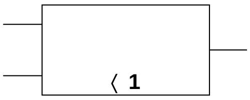
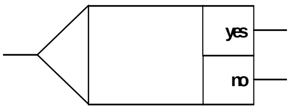
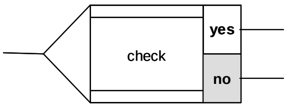
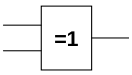
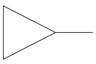
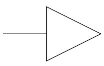
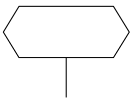


Fig. 10: Example for transition symbol

5.2.3 Summary of graphical symbols

An overview about the described graphical symbols for the modelling of the information flow is illustrated in Tab. 3. Finally it has be said that the symbols can be grouped into information and system components (1), actions among objects (2), linkages (3), branching and checks (4), exclusive OR-linkages (5), dividing the information flow into several graphs (6), and transitions (7).

Tab. 3: Graphical symbols in information flowchart

No	symbol	meaning	remark
1		information	Information and data objects, which occur as input, intermediate result or output within information flow
		system component	components (e.g. sensors, software, operator) as input variable or influence quantity of information quality
		comment	comment
2		line of action	Line of action for connecting information and system components with applications.
3		AND-linkage	Application, that connects input information by logical AND.
		OR-linkage	Application, that connects input information by logical OR.
4		single branching	Application, that divides the information flow into two paths in dependence of a query.
		single check	Application, that has a requirement to the input. In subject to the two possibilities “fulfilled” and “not fulfilled” the application can bring two possible disjunctive system states. This symbol divides the information flow into two paths. Two or more inputs will be connected by AND.
		arrow	The arrow symbol will only be used after single check, if data do not fulfil the requirements. This symbol is used before a new process, modelled by a system component
5		exclusive OR-linkage	For the connection of disjunctive states. This symbol will only be used after single branching and single checks.
6		carry input	For the break or continuation of the information flow in another graph.
		carry output	
7		transition symbol	Influence of one quality characteristics to the same or another quality characteristic. This symbol exists only related to the applications „linkage”, “branching” and “check”.

5.3 Dependencies among quality characteristics in information flow

5.3.1 General remarks on dependencies

Among the quality characteristics exist dependencies, which depend on the application types (see chapter 5.3.2). Dependencies may transfer a quality characteristic failure in the input to the same or to another quality characteristic failure in the output:

- Normally a quality characteristic failure in the input is transferred to the same quality characteristic failure in the output. E.g. the processing of incorrect input information is transferred to incorrect output information.
- In the exceptional case a quality characteristic failure in the input is transferred to another quality characteristic failure in the output. E.g. the processing of out-of-date input information is transferred to incorrect output information.

5.3.2 Application types

For dealing with dependencies and transfer of the quality characteristics the following application types will be defined:

- Processing
- branching
- check
- Usage
- Transfer
- Storage

These application types and their influences to the dependencies among quality characteristics are explained in the following.

Processing

The input information will be converted into an output information by using AND-linkages. An example for this linkage is illustrated in Fig. 11. This most frequently used application type, has several dependencies for quality characteristics. On the one side out-of-date input information is transferred to incorrect and / or unavailable information.

The influence from up-to-dateness on correctness can be explained by the following example: The geometry of a road element has changed, but there is no current maintenance of the database. The consequence of this out-of-date database is, that the calculated length of the road element will be wrong. So the out-of-date road element is also incorrect. On the other side the output information is up-to-date, because it is produced in the moment.

The influence from up-to-dateness on availability can be explained by the following example: A new road element was built-up, but there is no current maintenance of the database. The consequence of this out-of-date database is, that the length of the road element can not be calculated. In the result the length is not available.

These described influences from up-to-dateness on correctness and availability are valid in the same way for the influence from consistency.

The influence from completeness on correctness can be explained by the following example: In “Scenario A” an additional road element, which do not exists in reality, is stored in the database. The consequence of this commission is, that a routing in this street network can be incorrect. In the result the route is incorrect.

The influence from completeness on availability can be explained by the following example: In "Scenario B" road elements are missing in the database. The consequence of this omission is, that a routing in this street network can not be calculated. In the result the route is unavailable.

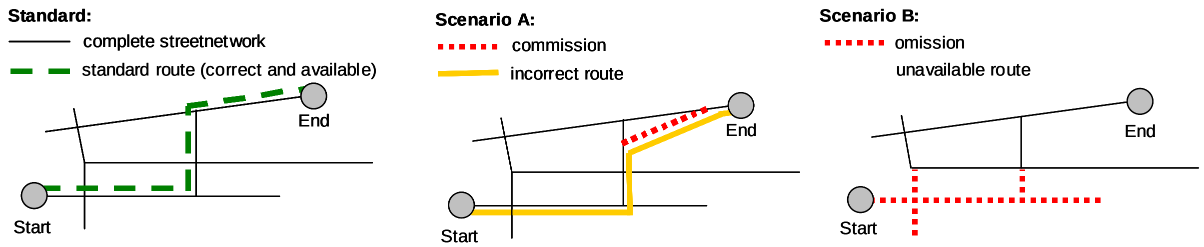


Fig. 11: Examples for influence of completeness

Branching

The information flow can be split into different information paths, depending on particular criteria. One possible criterion is the availability of the input information. Thereby the YES-path describes the case of available information and the NO-path the probability, that no information is available. The fact that the information is available with 100 % is implicit given by the calculation formulas (see chapter 5.4). An implicit influence between different quality characteristics, which has to be modelled by an transition symbol, does not exist.

Check

The information flow can be split into different information paths, depending on the fulfilment of a quality characteristic (with except for availability and accuracy). Additionally an information can be eliminated, if it do not fulfil a special quality requirement.

By using the application type "check" all quality characteristics have influences on availability, which will decrease. The reason for this is the elimination of an information, that do not fulfil a specific quality requirement. The eliminated data are not available anymore. All the influences between the quality characteristics will be taken into account by the calculation formulars (see chapter 5.4).

Usage

There is only one input and the same output. The application type "usage" models the influence of the time (up-to-dateness) to the information. E.g. by using a digital map for a time the quality characteristic "up-to-dateness" will be degraded. This influence is modelled by using the transition symbol (see Fig. 10).

Transfer

The application type "transfer" is used to model the transfer of information. Here only one input information is transferred the same output information, but the quality is changed through the influence of external system components. E.g. the data is transferred from the server to the car by using GSM (Global System for Mobile Communications). The availability (AV) of GSM has influence to the up-to-dateness (UP) of the data in the car (see Fig. 12). Consequently the influence of hardware components is very important for the transfer.

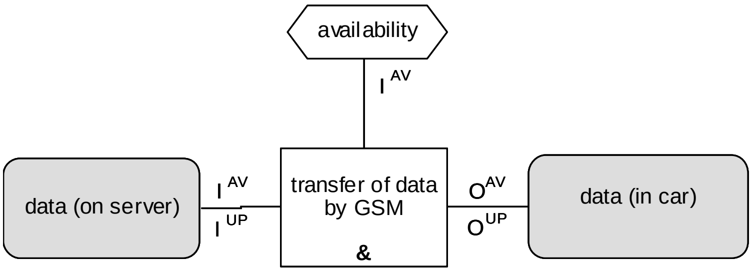


Fig. 12: Example for transfer

Storage

The application type "storage" should be used, when temporary data are stored in a permanent storage , with the intention to use of the data for longer periods (e.g. CD-Rom of street data). Therefore a transition symbol is used, which includes the dependence of the quality characteristic "availability" (AV) to the quality characteristic "completeness" (CM). The transition from availability to completeness is motivated by the reason, that there does not exist any time critical aspect, when a database is stored for long time using.

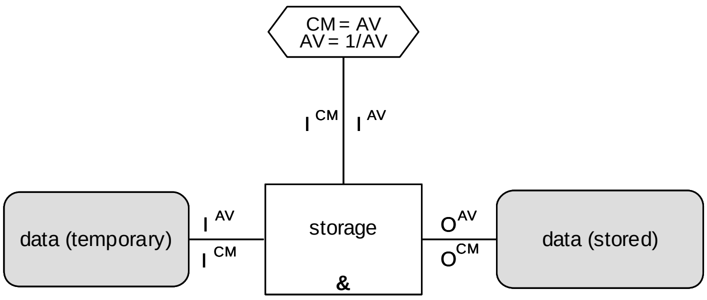


Fig. 13: Example for storage

5.3.3 Summary of dependencies

An overview about the described dependencies among quality characteristics is illustrated in Fig. 14. Finally it has to be said that:

- Each input quality characteristic failure is transferred to the same output quality characteristic failure.
- The quality characteristic "accuracy" has no influence on other quality characteristics. Therefore the accuracy can be dealt separately.
- By using the application type "single check" each quality characteristic (excluding accuracy) has an influence on the quality characteristic "availability". The dependencies of

the application type "single check" are not illustrated in Fig. 14 to allow a better clarity for the other dependencies.

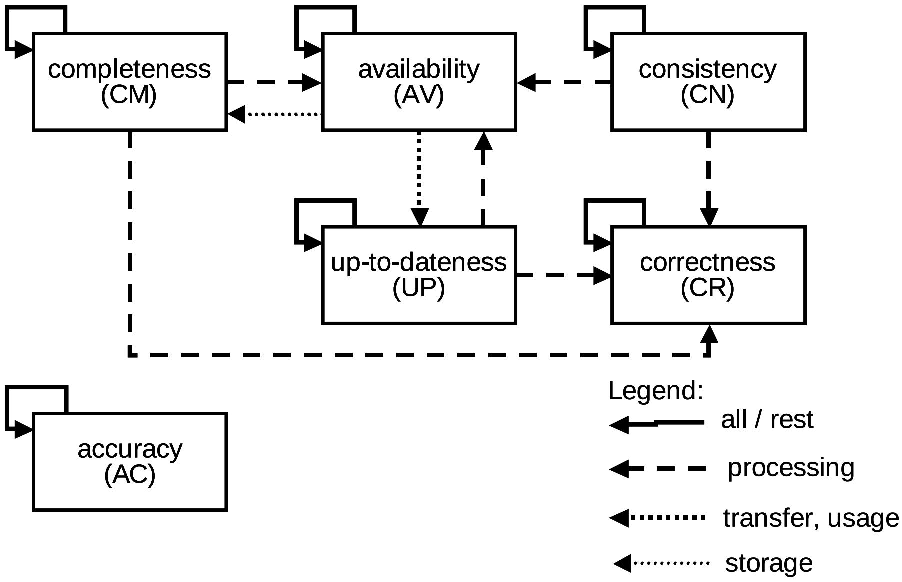


Fig. 14: Dependencies among quality characteristics in the information flow

5.4 Computing part for evaluating the information flowchart

AND

The AND-linkage is the most important linkage. The quality requirement at the output is only fulfilled if each input information fulfil the quality requirement. In the reell algebraic notation this corresponds to a multiplication of the quality values of the *n* input information I_i :

$$O^{QC(q)} = \prod_{i=1}^{n} I_i^{QC(q)} \quad \text{for } q \in [1..5] \tag{5-3}$$

OR

In the field of reliability analysis the OR-linkage is used for the description of technical availability of an active redundance. For the availability, which describe the existence of information the expression of the OR-linkage can be used directly:

$$O^{QC(q)} = 1 - \prod_{i=1}^{n} \left(1 - I_i^{QC(q)}\right) \quad \text{for } q \in [1] \tag{5-4}$$

For any other quality characteristics the probability of fulfilment of the quality requirements is given by the weighted average, whereas the weighting depends on the availability of the n input information:

$$O^{QC(q)} = \sum_{i=1}^{n} \left(\frac{I_i^{\mathrm{VV}}}{\sum_{j=1}^{n} I_j^{\mathrm{VV}}} \cdot I_i^{QC(q)} \right) \quad \text{for } q \in [2..5] \tag{5-5}$$

The use of the weighted average bases on the fact, that an information with a higher availability is used with higher probability then an information with a less availability.

Single branching

The single branching is used to divide the information flow in dependence of specified criteria. Following criteria can be applied:

- Dividing dependent on probability of availability of information
- Dividing dependent on an information value
- Dividing dependent on different system status.

In all cases the single branching divides the information flow in two disjunct system statuses. The probability of the system statuses is affected by the criteria.

If the availability of information is used as criteria, the YES-path gets the probability of of availability and the NO-path gets the complementary probability of the system status:

$$p_{YES} = I^{AV} \quad \text{and} \quad p_{NO} = 1 - p_{YES} = 1 - I^{AC} \tag{5-6}$$

On the NO-path no information is given out. Consequently no quality vector is given, but the probability of system status for no information.

For the quality values in the YES-path the following equations are valid:

$$\begin{aligned} O_{YES}^{AV} &= p_{YES} \cdot 1 \\ O_{YES}^{QC(q)} &= p_{YES} \cdot I^{QC(q)} \quad \text{for } q \in [2..5] \end{aligned} \tag{5-7}$$

The equations take into account the probability of the system status, for the case that the input information is available.

If the information flow should be divided into several path depends on different information values (e.g. road class 1, 2, 3 and 4) or system status (e.g. speed alert system operating in urban area or non-urban area) the probability of the outputs will be given by the chosen branching criteria. In these cases a divide into more than two paths are possible (**multiple branching**). For example we can divide the road data in dependence of the road class. The probability of each path will be calculated from the proportion for the road class (e.g. class 1: 8 %, class 2: 12 %, class 3: 23 % and class 4: 57 %).

Single check

This application is used for modelling quality checks in the information flow. As check criteria all quality characteristics are possible except for accuracy (treated using error propagation) and availability (treated using single branching). In the following explanations we will use the correctness as check criteria.

It is presumed that an information, which is detected as erroneous will be eliminated from the information flow. Consequently this induces a decrease of the availability of the information. On the other hand, the quality belonging to the checked quality characteristic will arise.

Secondly, it is assumed that it is not possible to detect every error. This aspect will be taken into account by means of the check quality. If we define a check quality $C_{YES} = 0{,}9$ we assume that 90 % of the incorrect information will be detected and eliminated of the information flow. The fraction of $C_{NO} = 1 - C_{YES}$ of the incorrect information will not be identified, and consequently not eliminated.

Because of the elimination of information, which are detected as erroneous, there is no information in the NO-path. In dependence of the availability and probability of the checked quality characteristic, we can calculate the probability for the NO-path:

$$p_{NO} = 1 - I^{AV} \cdot \left(C_{YES}^{QM(k)} \cdot I^{QM(k)} + C_{NO}^{QM(k)} \right) \quad \text{for } k \in [2..5] \qquad (5\text{-}8)$$

At this the variable k is the checked for quality characteristic. The probability of system status of the YES-path is similar to the complementary probability of the NO-path:

$$p_{YES} = 1 - p_{NO} = I^{AV} \cdot \left(C_{YES}^{QM(k)} \cdot I^{QM(k)} + C_{NO}^{QM(k)} \right) \quad \text{for } k \in [2..5] \qquad (5\text{-}9)$$

This means that we will use the YES-path if the input information is available and the check operation has not identified an error.

For the quality values in the YES-path count the following formulas, which take the probability of system status into account:

$$\begin{aligned} O_{YES}^{AV} &= p_{YES} \cdot 1 \\ O_{YES}^{QC(k)} &= p_{YES} \cdot \left(C_{NO}^{QC(k)} \cdot \sum_{i=1}^{n} I_i^{QC(k)} + C_{YES}^{QC(k)} \right) && \text{for } k = q \\ O_{YES}^{QC(q)} &= p_{YES} \cdot I^{QC(q)} && \text{for } k \neq q \\ &\text{and with } q, k \in [2..5] \end{aligned} \qquad (5\text{-}10)$$

The probability of the system status depends on the availability of the information and the probability of fulfilment the requirement of quality characteristics for which the check will be carried out. Because different probabilities of system status for checking different quality characteristics existent, it is only possible to check one criteria at a time. If different characteristics have to be checked, this has to be modelled as a sequence.

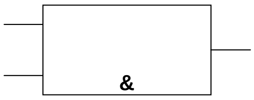
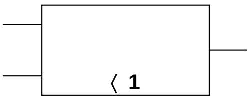
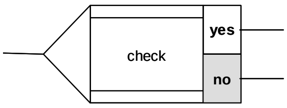
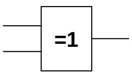
Another point which has to be taken into account is the fact that in many cases it is not feasible to decide, whether an error is a result from a failure of correctness, up-to-dateness or consistency. The check will only detect, that a gross error exist. Accordingly it is convenient to model a transit of the quality values from up-to-dateness and consistency to the correctness by using the transit symbol.

XOR

The linkage XOR is applied to disjunct conditions as they occur due to ramifications and checks. Following the rules of Boolean algebra the probability at the output is calculated by summation of the probabilities of the input. This counts for all quality characteristics.

$$O^{QC(q)} = \sum_{i=1}^{n} I_i^{QC(q)} \quad \text{for } 1 \le q \le 5 \tag{5-11}$$

Tab. 4: Formulas for the evaluation of the information flowchart

Symbol	Formula
AND-linkage  	$O^{QC(q)} = \prod_{i=1}^{n} I_i^{QC(q)}$ for $q \in [1..5]$
OR-linkage  	$O^{QC(q)} = 1 - \prod_{i=1}^{n} \left(1 - I_i^{QC(q)}\right)$ for $q \in [1]$ $O^{QC(q)} = \sum_{i=1}^{n} \left(\frac{I_i^{VV}}{\sum_{j=1}^{n} I_j^{VV}} \cdot I_i^{QC(q)} \right)$ for $q \in [2..5]$
Single branching  	$O_{YES}^{AV} = p_{YES} \cdot 1$ $O_{YES}^{QC(q)} = p_{YES} \cdot I^{QC(q)}$ for $q \in [2..5]$ with $p_{YES} = I^{AV}$ and $p_{NO} = 1 - p_{YES} = 1 - I^{AC}$
Single check  	$O_{YES}^{AV} = p_{YES} \cdot 1$ $O_{YES}^{QC(k)} = p_{YES} \cdot \left(C_{NO}^{QC(k)} \cdot \sum_{i=1}^{n} I_i^{QC(k)} + C_{YES}^{QC(k)}\right)$ for $k = q$ and $q,k \in [2..5]$ $O_{YES}^{QC(q)} = p_{YES} \cdot I^{QC(q)}$ for $k \ne q$ and $q,k \in [2..5]$ with $p_{NO} = 1 - I^{AV} \cdot \left(C_{YES}^{QM(k)} \cdot I^{QM(k)} + C_{NO}^{QM(k)}\right)$ for $k \in [2..5]$ $p_{YES} = 1 - p_{NO} = I^{AV} \cdot \left(C_{YES}^{QM(k)} \cdot I^{QM(k)} + C_{NO}^{QM(k)}\right)$ for $k \in [2..5]$
Exclusive OR- linkage  	$O^{QC(q)} = \sum_{i=1}^{n} I_i^{QC(q)}$ for $1 \le q \le 5$

5.5 Example speed alert

Considering as example of a speed alert system the application of the analysing procedure will be illustrated. Within the scope of the example several simplifications will be undertaken for a better and easier understanding. Firstly, the consideration will be limited on the quality characteristics availability, completeness and correctness. This results in a quality tupel for this example as follows:

$$I_{Data} = \left[I_{Data}^{AV}, I_{Data}^{CM}, I_{Data}^{CR}\right] \tag{5-12}$$

Secondly, a very simple and abstract information flow for the system modelling is chosen.

For the system two different system versions will be carried out:

- information flow without a quality check
- information with a check of completeness and correctness as a quality assurance measure.

Both system designs use a data source with the road network and a data source with speed traffic signs. By means of a matching and merging procedure the two data sources are linked together. The final step of the content provider is the storage of the geographic data in an exchange format. At this step a dependency between two quality characteristics occur: A failure of availability results in failure of completeness. For using the calculation rules of the AND-linkage the completeness of the transition symbol has to equate with the availability of the map:

$$I_{Trans}^{CM} = I_{Map}^{AV} \tag{5-13}$$

As the fault of existence of information is accordingly described by the completeness, the availability of the information is fulfilled. For this reason the availability of transition symbol is equal to the reciprocal of the availability of the input information “map”:

$$I_{Trans}^{AV} = 1 / I_{Map}^{AV} \tag{5-14}$$

The availability value of the output information of “storage” is consequently 1.0.

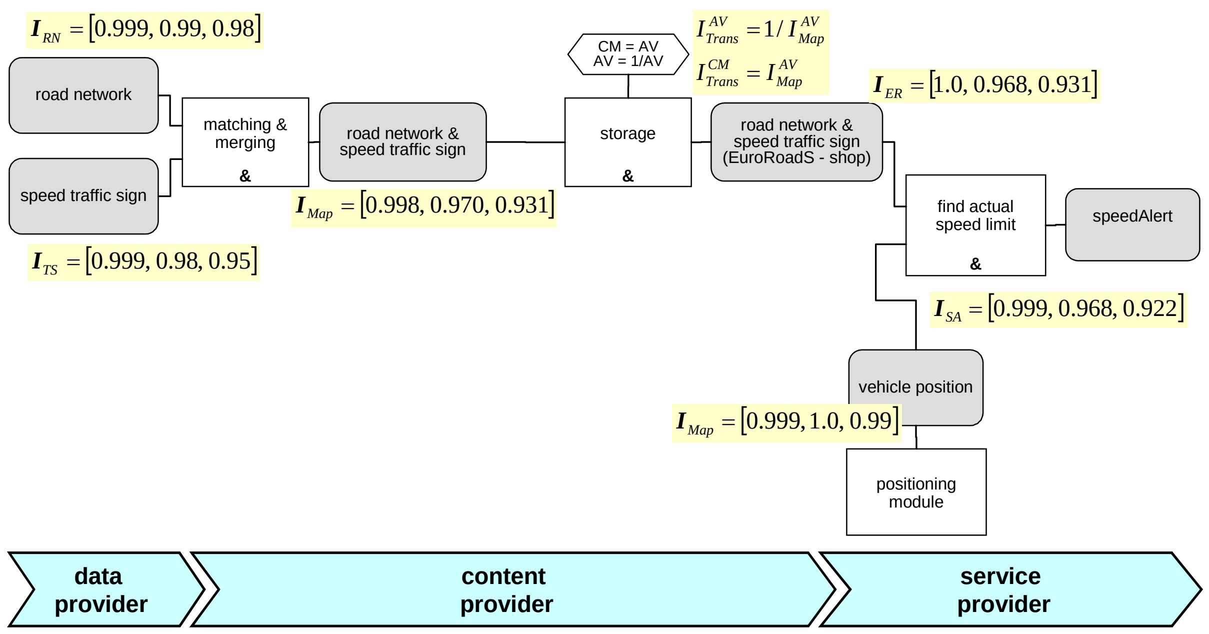


Fig. 15: Information providing process without quality check

For the first version, the content provider can deliver the database of the speed alert system with a probability of completeness of 96.8 % and of correctness of 93.1 % (see. Fig. 15).

For the second version, where a quality check is carried out, the database have a significant increased quality concerning the two error types completeness and correctness. In the result the database exhibit a probability of completeness of 99.3 % and correctness of 98.4 % (see Fig. 16).

The modelling of the quality check procedure take into account, that not every error in the database can detected as well as the fact that the manual reworking is not completely error-free.

The increased quality of the version two is transmitted in the quality of the speed alert of the driver assistance system. Without quality check of the digital map, a rate of completeness resp. correctness is expected of 96.8 % resp. 92.2 %. By using a digital map, which contains a quality check before delivery, the driver assistance belonging speed limitation can be effected with a serious increased quality of 99.3 % resp. 97.4 %.

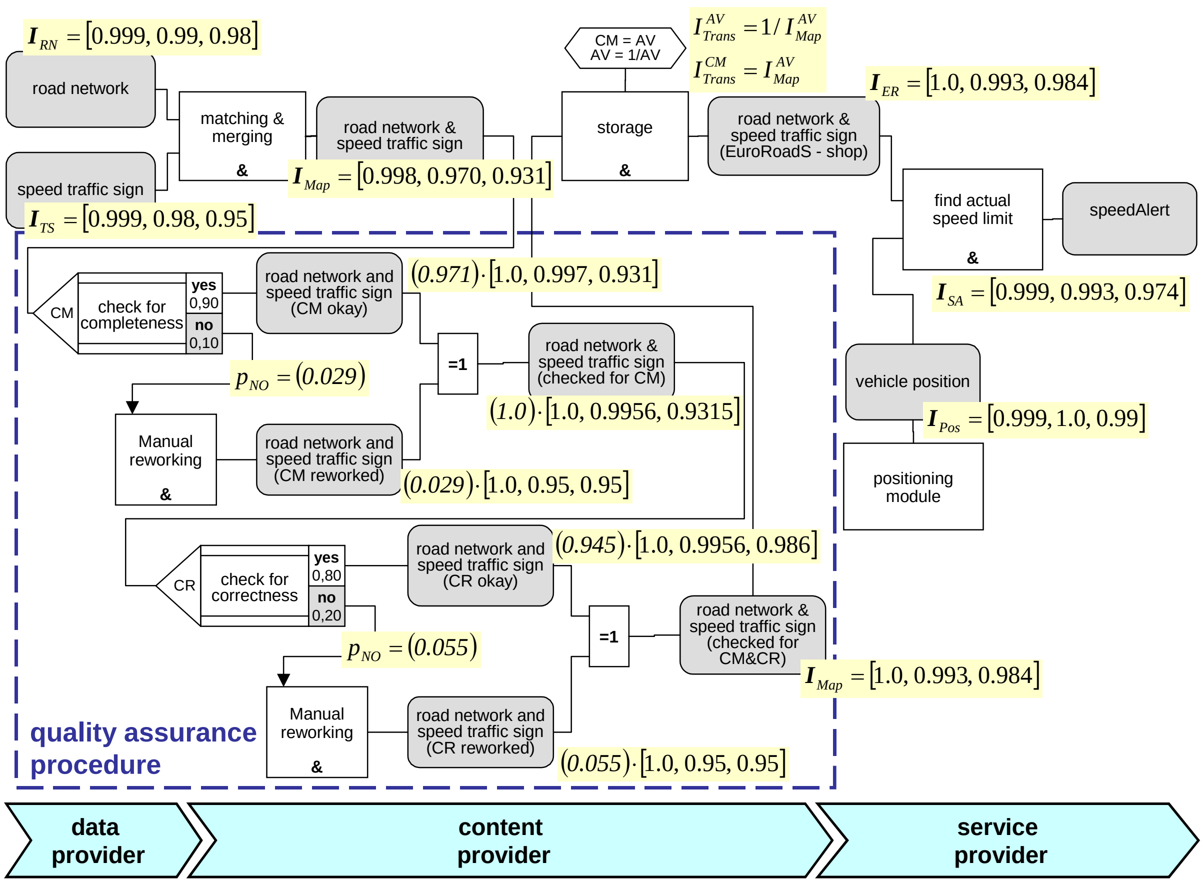


Fig. 16: Information providing process with quality assurance measures

6 Description and modelling of information quality of data providing processes

The following description of processes of data providing is based on current literature (Bill 1999a, Chen et. al 2001, Clarke 2001, Hake et. al 2002, Longley et. al 2001) and own experiences. Based on this research the information flow of the different providing processes, as typically occuring, were modelled in a general way by using the developed method (see chapter 5). All variations of these processes can in practice be modelled by the method. Also a more detailed modelling is possible.

For a better overview the identified functions within data providing processes are structured in a function tree (see Fig. 17).

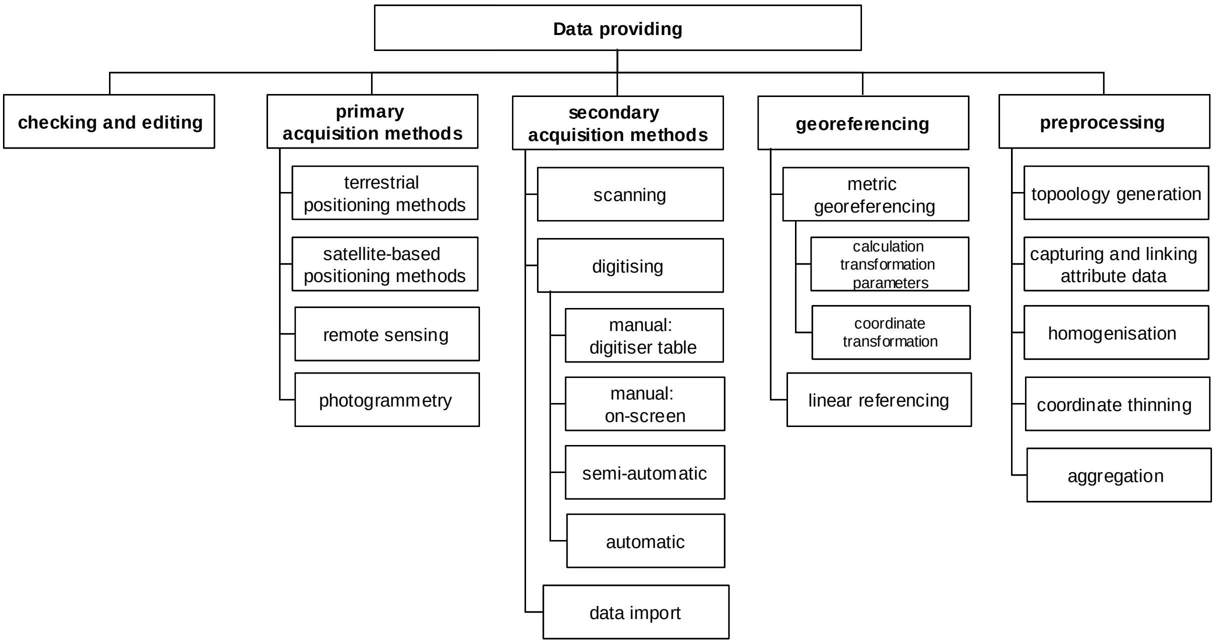


Fig. 17: Function tree of identified functions within data providing processes

The quality within the data providing processes depends generally of input data, hardware, software, and operator. The following general remarks regarding input information and system components give the framework for the description and modeling in this chapter:

- *Hardware*: It is a fact, that the use of high-grade hardware and expert implementation and maintenance reduces hardware failures and errors. In the following the hardware will not described explicitly, because it is assumed that the used hardware is functioning. The hardware is only be described explicitly, if its components are very important for the data providing process resp. providing process step.
- *Software*: It is a fact, that the use of professional and debugged software and software components reduces failures and errors. In the following the software will not be described explicitly, because it is assumed that the used software is functioning. A software will be described explicitly, when its algorithm is very complex and contains fuzziness in data processing (e.g. raster-vector-conversion, automatic error detection, or coordinate thinning). In all these cases the used and implemented algorithm in connection with defined limit value has an important influence on information quality.
- *Operator*: It is assumed that specialised staff are employed for the task. Thereby the continuously advanced training is a crucial point for the quality of the operator. On the one side the operator will not be described explicitly if it is an automated process. On the other side the operator will be described explicitly, if it is a manual process and accordingly human errors can influence the quality of information provision.

Finally, if one of these technical and human components (hardware, software, operator) have no significant impact on the quality result, the description of these components can be omitted.

In the following examples the data providing processes are described by using the information flowchart. These processes can also be evaluated by using the computing procedure, based on probabilistic methods.

6.1 Error check and editing

The check and the reworking (editing) is frequently used for other procedures, like checks for topology generation, linear referencing, capturing attribute data et al.

The dataset has to be checked for *slivers, gaps, spikes, inversions, lines that do not end, unsnapped nodes* or other errors in the dataset. There exists on the one side slivers, which occur when the adjacent polygons apparently overlap. On the other side gaps occur when there is adjacent space among adjacent polygons. These slivers and gaps may be generated if two maps are merged together. Thus either manual or automatic correcting processes are necessary for detection of slivers, gaps and other blunders. The reached quality of this processing step depends in particular on the basic data, the rate of error detection by the software (automatic error detection) and / or by the operator (manual error detection), and the quality of the reworking process.

Following the check the rework is often manually. I.e. the geometry, topology and attributes of datasets are often rework by using editing tools. The geometry may be edited by supplementing, copying, deleting, moving, rotating, dividing lines, joining lines, altering form or by other functions. The topology can be edited by splitting, merging and other functions. Besides the attributes can be edited. The quality of editing data depends in particular on the data and the operator.

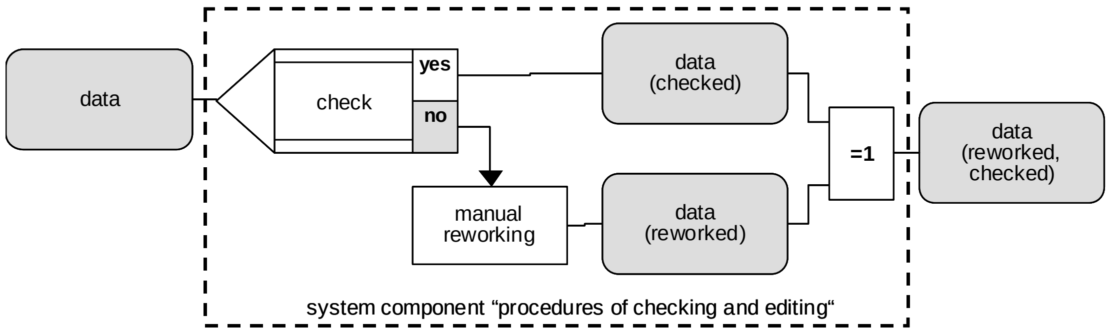


Fig. 18: Information flowchart for the procedures of checking and editing

The procedures of checking and editing are frequently used for other procedures (e.g. as part of the topology generation). In the following the procedures (see Fig. 18) will be simplified by the system component “procedures of checking and editing” as illustrated in Fig. 19.

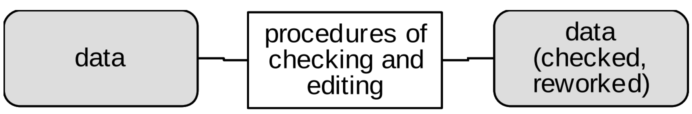


Fig. 19: Simplified information flowchart for the procedures of checking and editing

This example illustrates the ability of a modular structuring and modeling of the information system resp. the information process as well as the ability of composition and decomposition.

6.2 Primary methods of data acquisition

The primary methods of data acquisition are those acquiring the data from the object itself or from its unchanged image. This includes terrestrial and satellite-based positioning methods as well as remote sensing and photogrammetry. The quality of the acquired data are depending on these acquisition methods. For example in Tab. 5 the accuracy of different data acquisition methods are described. A detailed definition and modelling of separate steps of primary acquisition will not be given in these document, because the project EuroRoadS focuses on the steps of processing, which follows the acquisition. Nevertheless it would be possible to model the different steps of data acquisition by the method described. For more details about the primary methods of data acquisition is referred to the literature (Chen & Lee 2001).

Tab. 5: Characteristics of data acquisition methods (own compilation regarding to Bill 1999)

	elements	**accuracy**	**applications**
terrestrial measurement	point, line	cm – dm	local
GPS-positioning	point, line	cm – m	local to global
remote sensing	area	> 10 m	regional to global
stereo photogrammetry	point, line	$1 * 10^{-5} * s_p$	local to regional
digitising	point, line, area	$1 * 10^{-4} * s_m$	local to global
		s_p = scale picture s_m = scale map	

6.3 Secondary methods of data acquisition

The secondary methods of data acquisition are those acquiring data from existing paper maps and digital geographic data. This includes the methods of scanning and digitising as well as data import from existing data sources.

The precondition for scanning and digitising is the georeferencing of the basic map (see chapter 6.4.1).

6.3.1 Scanning

The raster data capture by drum scanner or flatbed scanner take place as follows: An analogue map which is placed on the flatbed and accordingly on the rotating drum will be scanned successively in lines. The quality depends in particular on the analogue map, and the hardware (detector et al.).

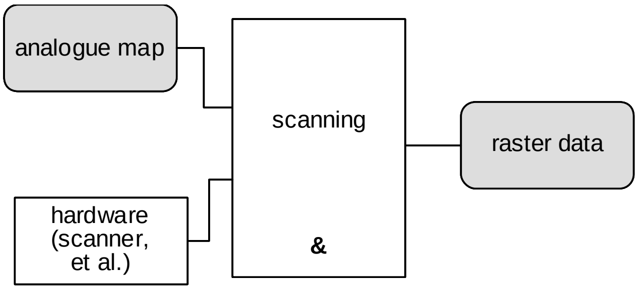


Fig. 20: Information flowchart for scanning

The procedure of scanning can be the first step for on-screen-digitising, semi-automatic vectorisation, and automatic digitising. In the following the procedure of scanning (see Fig. 20) will be simplified by the system component “scanning” as illustrated in Fig. 21.

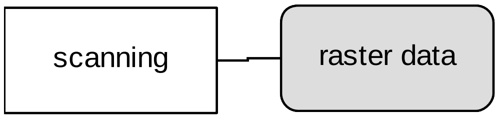


Fig. 21: Simplified information flowchart for scanning

6.3.2 Digitising

6.3.2.1 Manual digitising by digitiser table

The manual digitising by *digitiser table* take place as follows: An analogue map is placed on a flat tablet, and an operator follows the map features by using a cursor. The locations of features on the map are sent back to the computer every time the operator of the digitising table presses the button. Stream mode digitising partially automates this process by instructing the digitiser control software automatically to collect vertices every time a distance or time threshold is crossed. The quality depends in particular on the analogue map, the hardware (digitising table, cursor, et al.), and the operator.

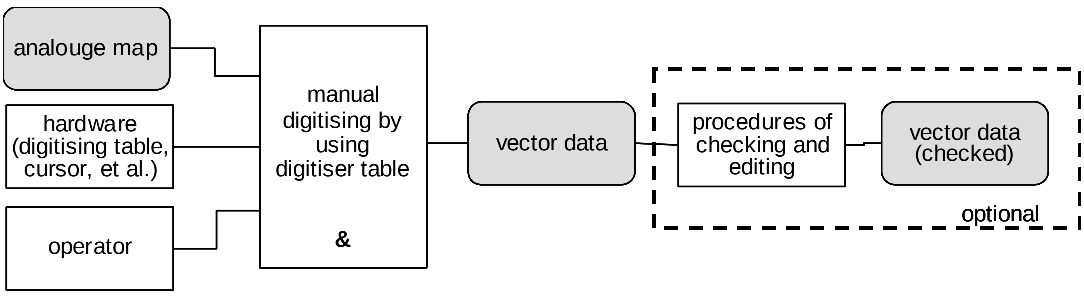


Fig. 22: Information flowchart for manual digitising by using digitiser table

6.3.2.2 On-screen-digitising

Another procedure is the *on-screen-digitising*. For this a scanned or imported digital map is digitised on the computer screen. The quality depends in particular on the raster data and the operator.

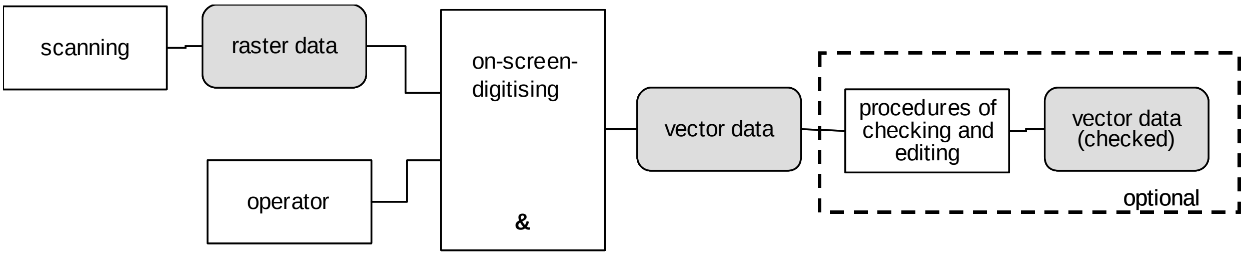


Fig. 23: Information flowchart for on-screen-digitising

6.3.2.3 Semi-automatic vectorisation

The *interactive vectorisation* (also called semi-automatic vectorisation, line following, or tracing) take place as follows: The operator snaps the cursor to the start point, the end point, the cross-over point and then he indicates the direction for line following. Then the line will be followed by using software. Therefore the quality depends in particular on the raster data, the software (tracing modus), and the operator.

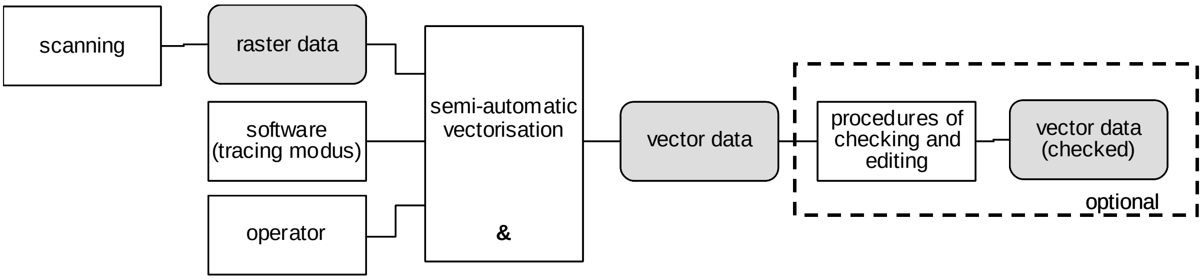


Fig. 24: Information flowchart for semi-automatic vectorisation

6.3.2.4 Automated digitising

The *automated digitising* consists of raster-vector-conversion (batch mode), and manual reworking. The first step, the batch vectorisation takes an entire raster file and converts it to vector objects in a single operation. Vector objects are created by software algorithms that build simple line strings (spaghetti data, unstructured vector data) from the original pixel values. Problems of the raster-vector-conversion can be round off edge, junction bridge, stubble, junction shifting, island. For the correction of these blunders interactive reworking steps are necessary, i.e. a post-vectorisation editing (manual reworking) is required to clean up errors.

Thus the quality depends in particular on the raster data, the software for conversion and optional on the procedures of checking and editing (see 6.1).

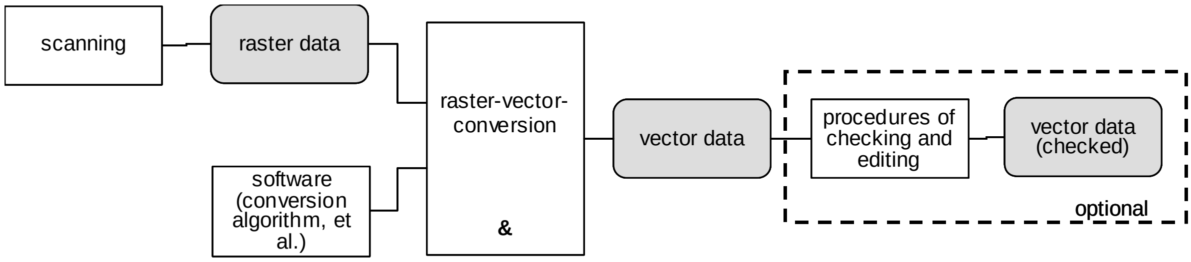


Fig. 25: Information flowchart for automatic digitising

Finally an overview about the different possible working steps of digitising is given in Fig. 26.

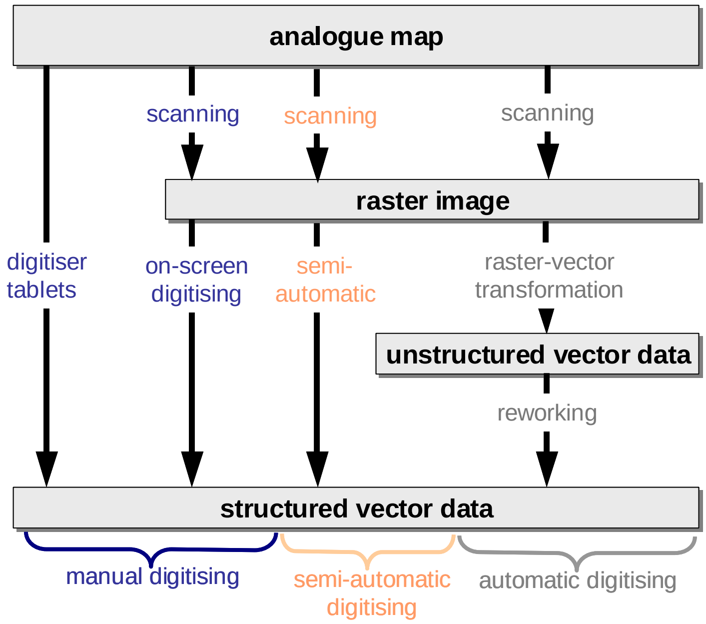


Fig. 26: Overview about the working steps of digitising

6.3.3 Data import

Besides data can be obtained from external sources. One of the biggest problems using data is that it may be encoded in many different formats. Direct read support is only be provided for simple product-oriented formats. Complex formats were designed for exchange purposes and require more advanced processing before they may be viewed. If the data can not be read by the target GIS software, then the data have to be converted. Therefore a geographic data translation software must address both syntactic and semantic translation issues.

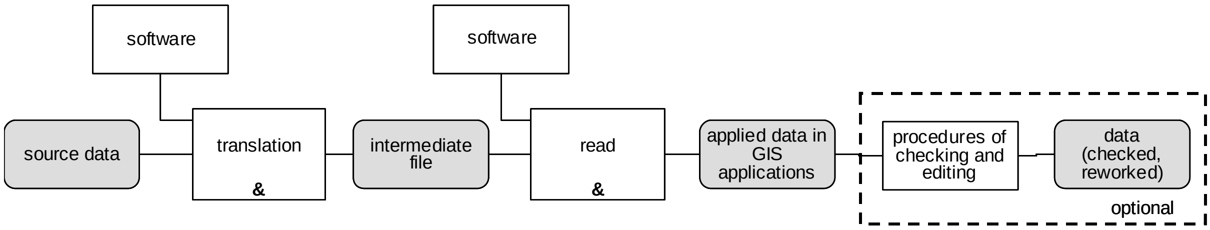


Fig. 27: Information flowchart for data import

Possible sources of quality interference may be corrupted media, incomplete data files (completeness), wrong versions of translators, different interpretations of a format specifications (consistency), and user errors.

6.4 Georeferencing

The techniques for specifying location are described by the verbs *to georeference*, *to geolocate* and *to geocode*. It has to distinguish among the types of metric georeferencing and linear referencing.

6.4.1 Metric georeferencing

The metric georeference is the one which assign geographic data to a defined coordinate system (e.g. WGS 84, Gauß-Krüger) including the coordinate values (e.g. latitude and longitude). The first step is to calculate the transformation parameters and then to transform the dataset.

6.4.1.1 Calculation transformation parameters

The quality of the transformation parameters depends in particular on the reference points, and optional on the procedures of checking and editing.

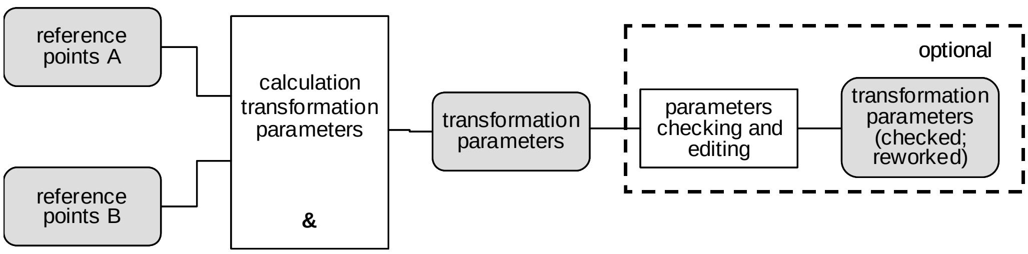


Fig. 28: Information flowchart for calculation transformation parameters

6.4.1.2 Coordinate transformation

The quality of the Coordinate transformation depends in particular on the data, the transformation parameters, and optional on the procedures of checking and editing.

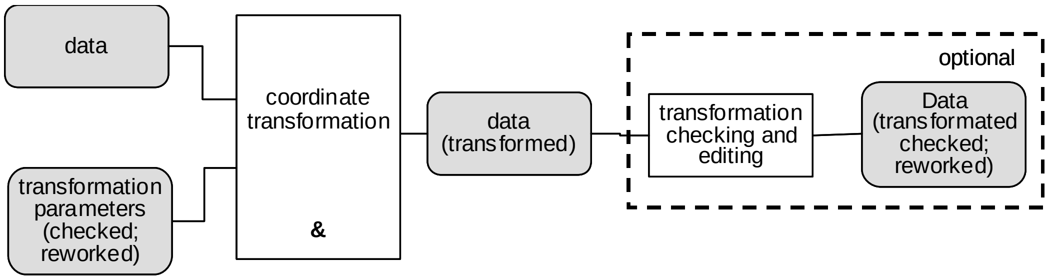


Fig. 29: Information flowchart for coordinate transformation

6.4.2 Linear refererencing

Another kind of georeferencing can be using a linear referencing system, whereby locations (e.g. signs, bridges or accidents) will be defined on a network by measuring distance from a defined point of reference along a defined path in the network.

The quality of linear referencing system depends in particular on the data, the software algorithm, and optional on the procedures of checking and editing.

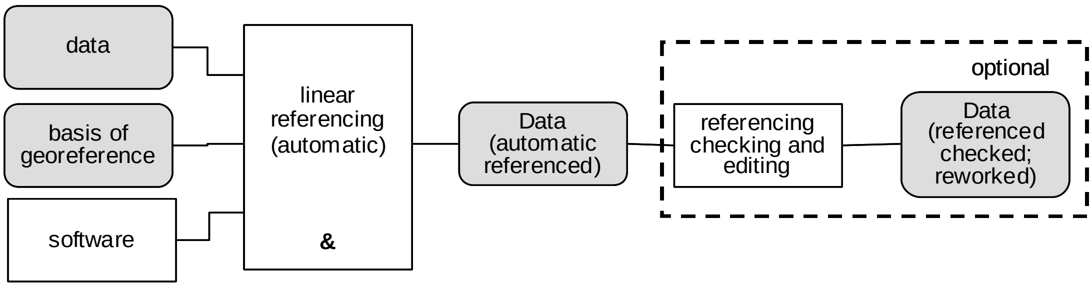


Fig. 30: Information flowchart for linear referencing system

6.5 Topology generation

The topology generation includes the processing steps from simple features (unstructured vector data) to topologic features (structured vector data) by using topological rules. The topology is used to validate the geometry of vector datasets (e.g. check that polygons are closed or that all lines in a network are joined together), and for certain types of operations (e.g. network tracing or tests of polygon adjacencies). Creating the topology is usually an iterative process because to resolve all data problems in an automated way is seldom possible.

The quality of topology generation depends in particular on the vector data, the operator and the procedures of check and editing.

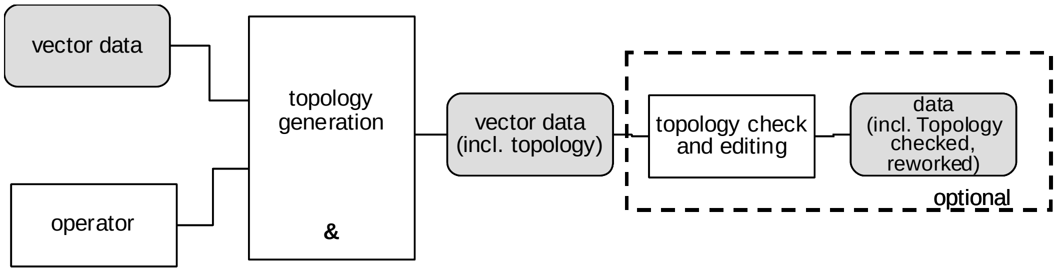


Fig. 31: Information flowchart for topology generation

6.6 Capturing and linking attribute data

Attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition or voice recognition. The most common method is direct keyboard data entry into a spreadsheet or database. For capturing attribute data a validation check is also necessary and maybe a manual reworking. An essential requirement is that object geometry and attributes can be linked together by using an identifier.

Besides metadata, as a special type of non-geometric data, are increasingly being collected. Some metadata may be deviated automatically by the GIS software (for example: length and area, extent of data layer, and count of features), but some have to be collected explicitly (for example: owner name, quality estimate, and original source).

The quality of capturing and linking attribute data depends in particular on the attribute source, the operator, the procedures of checking and editing, and finally the geometry and software for linking.

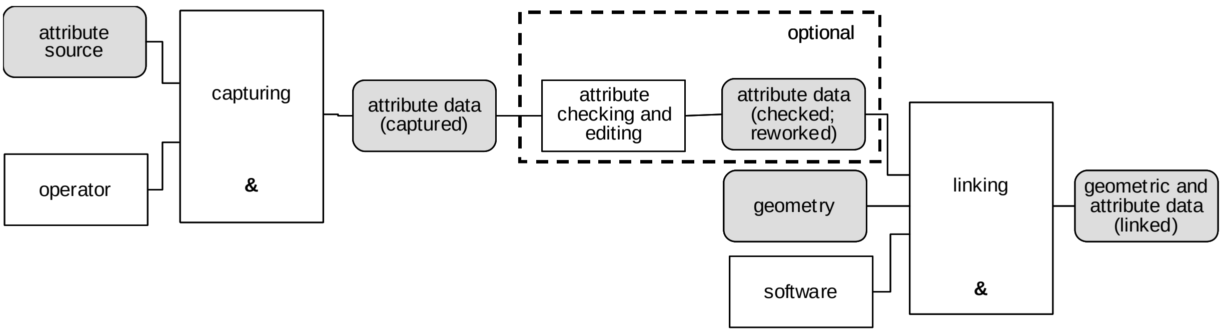


Fig. 32: Information flowchart for capturing and linking attribute data

6.7 Homogenisation

Geometric restrictions like right angle, parallelism, straight-line are necessary to upgrade the quality of digitising. That means for example that edges of a polygon have to be in a right angle; which can be realised by homogenisation.

The quality of homogenised data depends in particular on the basic data, the software, the operator, and optional on the procedures of checking and editing.

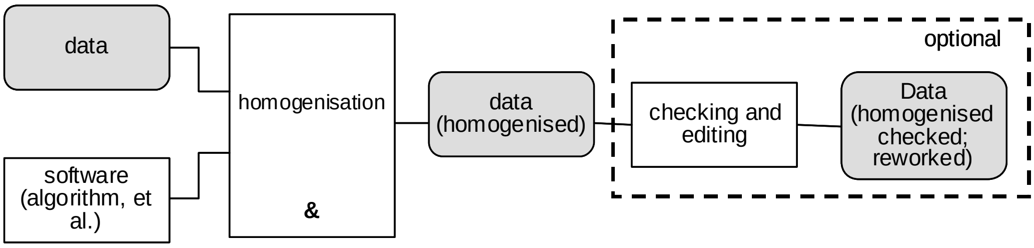


Fig. 33: Information flowchart for homogenisation

6.8 Coordinate thinning

In the basic map (e.g. scale 1:1000) are often more coordinates (data) than are needed to define a line or polygon in the derived map (e.g. scale 1:10000). Therefore coordinate thinning is necessary, which reduces the quantity of coordinate data by deleting redundant data. The simplest approach is to delete every n^{th} point (every 2^{nd} or 3^{rd}, etc. point). More advanced approaches are using the Douglas and Peuker algorithm or Reuman and Witkam algorithm.

The quality of coordinate thinning depends in particular on the basic data and the software algorithm, and optional on the procedures of checking and editing.

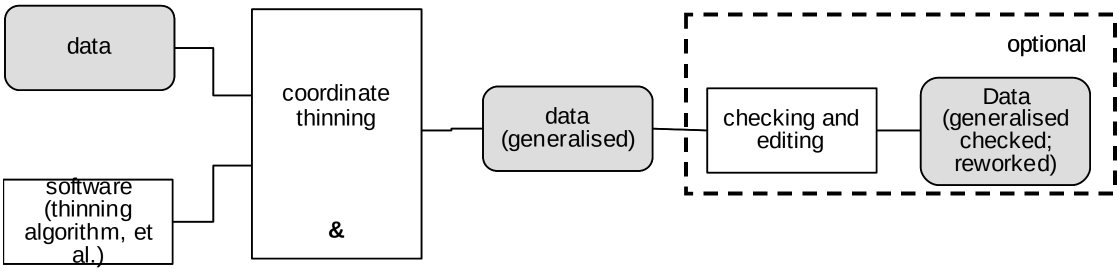


Fig. 34: information flowchart for coordinate thinning

6.9 Data aggregation

If the required information exists in more than one map source, different working steps of data aggregation may be necessary. Therefore a set of tools within GIS can exists for merging, dissolving or clipping multiple data sets in order to establish the required information. The level of automation of these tools can variegate in practice The function `*merging*` combines two or more adjacent layers into one layer that contains all their features (`edge matching`). The function `*dissolve*` aggregates features that have the same value for an attribute thereby simplifying data. By using the function `*clip*` a piece of one layer will be cut out like a cookie cutter.

The quality of data aggregation depends in particular on the data sets and the software, and optionally on the procedures of checking and editing.

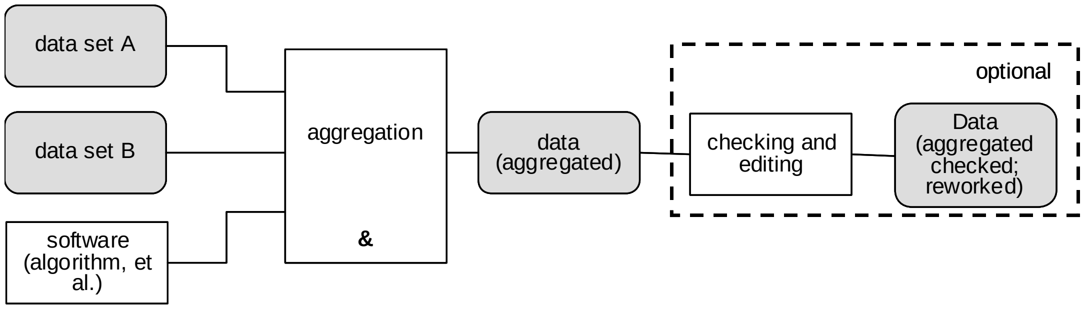


Fig. 35: Information flowchart for aggregation

6.10 Outlook: Data analysis and visualisation

The functions of data analysis and visualization are important for the service applications. E.g. the quality of routing analysis depends on the road network, the start and end point, and the software.

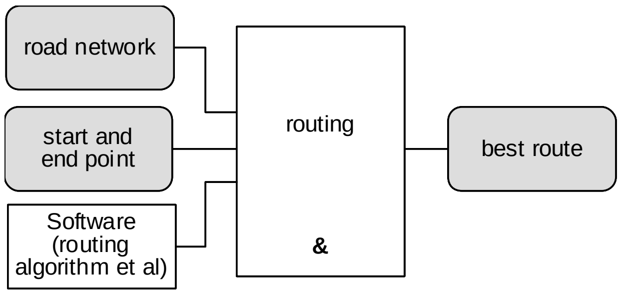


Fig. 36: information flowchart for routing

The display depends on the scale, the resolution, the hardware. A detailed definition and modelling of separate steps of data analysis and visualization will not be given in these document, because its focused on the steps of processing. The analysis and visualisation of service application are part of work package WP7 (Demonstrator).

7 Conclusion and Outlook

The aim of deliverable D2.3 was to specify a method to describe and evaluate information quality within geoinformation processes. For this analysing procedure the following requirement are postulated:

- Intuitive understandable modelling of the information flow in form of a graphical representation, which illustrates all processing steps and components with influence on information quality.
- Possibility of a modular structure of the information processes.
- Evaluation of the information flow by means of probability calculation.
- Usable for the analysis and evaluation of existing processes as well as development and design of future information processes.

- An interdisciplinary applicability of the analysing procedure so that every involved party can be participate and support the process.

To fulfil postulated requirements a method was derived from the reliability analysis method of mechanical engineering. The method includes the graphical information flowchart, by which the data providing processes can be represented by using defined graphical symbols. The method includes also a computing procedure, by which the quality can be analysed and evaluated by using probabilistic calculation.

Furthermore the specified method was used to model in a general way the information flow of the different providing processes, like they typically occur. At this the processes were structured into the functions of acquisition, processing and analysis as well as the included checking and editing procedures within the information chain.

Based on this modelling in a general way, described in the existing deliverable D2.3, the data providing processes for the EuroRoadS demonstrator will be modelled and evaluated in detail, being described in the following deliverable D7.5.

The described analysing procedure is a method for simulation quality assurance within geoinformation processes. The workflow of real processes can be described by the well known Deming wheel or cycle (PDCA), which was originally conceived by Walter Shewhart in 1930`s, and later adopted by W.Edward Deming. The model provides a framework for the improvement of a process or system, which include Plan, Do, Check and Act. These four components and the integration of EuroRoadS quality elements into the PDCA-cycle will be described finally:

Plan: Establish the objectives and processes necessary to deliver results in accordance to the specification. The required (product) quality can be described by the set of quality characteristics (D2.2) and formulated by the quality parameters. The analysing procedure (D2.3) can be used to design quality assurance measures. On basis of the model of the data providing process, quality lacks in the process can be identified and quality assurance measures can be integrated. Their impact on to the quality can be evaluated by the method.

Do: This phase contains implementation of measures in real data providing process.

Check: Monitoring and evaluating the processes and results against objectives and specifications and report the outcome. Therefor the evaluation methods for measuring quality parameter values (D2.5) can be used.

Act: After planning, implementing and checking, it has to be verified whether the required product quality is fulfilled.

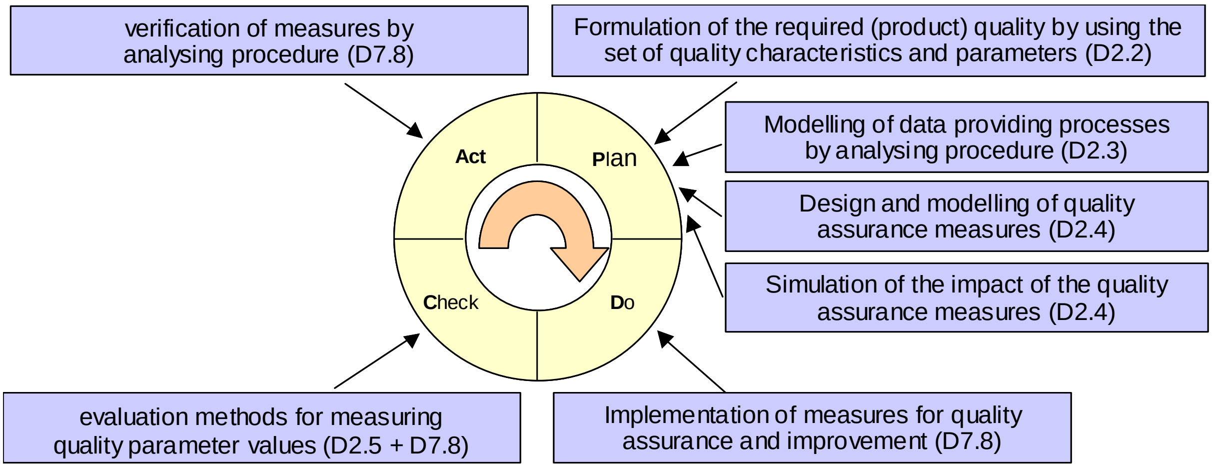


Fig. 37: Integration of EuroRoadS quality concept into the PDCA-cycle.

Annex A: Literature

BILL, RALF (1999): Grundlagen der Geo-Informationssysteme; Band 1 – Hardware, Software und Daten. Wichmann. Heidelberg.

BIROLINI, ALLESANDRO (1999): Reliability Engineering. Theory and Practice. Springer. Wien, New York.

BLÄSING, JÜRGEN P.; EICHE, DANIEL (2002): Workbook FMEA: Failure Mode & Effects Analysis. TQU-Verl. Ulm.

CHEN, YONG-QI; LEE YUK-CHEUNG (2001): Geographical data acquisition. Springer. Wien, New York.

CLARKE, KEITH C. (2001): Getting started with geographic information systems. Prentice Hall. Upper Saddle River.

HAKE, GÜNTER; GRÜNREICH, DIETMAR; MENG, LIQIU (2002): Kartographie. Walter de Gruyter. Berlin, New York.

LONGLEY, PAUL A.; GOODCHILD, MICHAEL F.; MAGUIRE, DAVID J.; RHIND, DAVID W. (2001): Geographic Information Systems and Science. Wiley. Chichester.

UNIVERSITY ROSTOCK (2004): Geoinformatik-Service. (*http://www.geoinformatik.uni-rostock.de*). December 2004. Rostock.

WILTSCHKO, THOMAS; KAUFMANN, THILO (2004): Quality frame for information. EuroRoadS project deliverable D2.2.

Annex B: Glossary

Term	Definition
quality concept	Concept to describe and evaluate quality of geoinformation within data providing processes.
quality model	Part of the quality concept that contains a fixed set of inherent quality characteristics and variable quality parameters to describe the quality of geoinformation. For a fixed set of inherent quality characteristics an adapted structure of the quality element and subelements of ISO 19 113 is necessary.
analysing procedure	Part of the quality concept that contains information flowchart and computing procedure to model processes. It can be used for evaluation of quality assurance measures within geoinformation processes. It is based on a probabilistic model.
information flowchart	Graphical part of the analysing procedure to describe and evaluate quality within information processes. Thereby the processes can be represented by using defined graphical symbols.
computing procedure	Analytical part of the analysing procedure to describe and evaluate quality within information processes. Thereby the quality can be analysed and evaluated by using probabilistic calculation.
information	Object of the analysing procedure that occur as input, intermediate result or output within information flow.
application	Object of the analysing procedure in which input information enter and output information goes out. The application is illustrated by a symbol used in the information flow graph and contains a calculation rule used in the computing part.
system component	Object of the analysing procedure that is used to model technical or human influences on the information quality.
single branching	Application that divides the information flow into two paths in dependence of a query.
AND-linkage	Application that connects input information by a logical AND.
OR-linkage	Application that connects input information by a logical OR.
single check	Application that can bring two possible disjunctive system states in regard to the two possibilities “fulfilled” and “not fulfilled”.
exclusive OR-linkage	Application that connects disjunctive states.
transition	Application that models the influence of one quality characteristics to the same or another quality characteristic.
application type	Types of the analysing procedure that have typical dependencies between quality characteristics.
processing	Application type that use “AND-linkage”.
branching	Application type that use “single branching” and split the information flow into different information paths, depending on particular criteria.
check	Application type that use “single check” and split the information flow into different information paths, depending on the fulfilment of a quality characteristics.
usage	Application type that use “transition” and models the influence of the time to the information.
transfer	Application type that use “transition” and models the influence of external system components to the information.
storage	Application type that use “transition” and models that temporary data are stored in a permanent storage with the intention to use it for a longer time.
data providing processes	Acquisition and processing methods as well as the included checking and editing procedures within the information chain.