Quality of Safety-Related and Real-Time Traffic Information Services

Quality package

EU EIP SA 4.1: Determining Quality of European ITS Services

Version: 2.0

Date: May 15, 2019

Co-financed by the European Union
Connecting Europe Facility

Document Information

Author

Name	Organisation
Risto Kulmala	Finnish Transport Agency, FI
Risto Öörni	VTT Technical Research Centre of Finland, FI
Tomi Laine	Strafica Ltd, FI
Peter Lubrich	BASt, DE
Anna Schirokoff	Finnish Transport Safety Agency, FI
Teun Hendriks	TNO, NL
Leif Rystrøm	Danish Road Directorate (DRD)

Distribution

Date	Version	Dissemination/Content
2018-02-26	1.1	Former version : Publication on EU EIP web site
2018-11-20	1.7	New version after validation phase - Activity 4.1 internal
2019-02-18	1.8	Activity 4.1 internal
2019-03-15	1.9	Activity 4.1 internal
2019-03-19	2.0	Activity 4.1 internal
2018-05-15	2.0	Publication on EU EIP web site

Preface

The European ITS Platform (EU EIP) has initiated a framework for the quality of Safety-Related Traffic Information (SRTI) and Real-Time Traffic Information (RTTI) services, as well as of their data contents.

This document presents the results in form of a “Quality Package”, containing quality-related definitions and concepts, proposed and agreed by EU EIP partners for the use in Europe. The results are based on evidence from conditions and operating environments in combination with the expert knowledge of the public and private stakeholders involved in the EU EIP quality work.

A previous version of this framework has been published in February 2018. Since then, EU EIP has continued working on the validation of the “Quality Package”. This validation has been based on new quality tests among several EU EIP partners. As a result, some changes to the quality criteria and quality levels have been proposed and incorporated.

This document contains the most-recent version of the “Quality Package”, considering updated and validated definitions and recommendations on Quality Criteria, Quality Requirements and Quality Assessment methods for SRTI and RTTI services.

The document is prepared as part of the sub-activity 4.1 of the EU EIP project.

T A B L E   O F   C O N T E N T S

1. Introduction

This document presents a framework regarding the quality of Safety-Related Traffic Information (SRTI) and Real-Time Traffic Information (RTTI) services, as well as of their data contents, to be used in Europe.

It has been elaborated by EU EIP sub-activity 4.1, considering expertise of European stakeholders and practitioners on this field. The work summarised here is backed up on years-long experience in quality frameworks, including an iterative validation of its contents and a continuous stakeholder consultation.

The use of such a quality framework serves the objectives of both the European Commission (EC) as well as the EU member states, because it relates to the ITS Directive 2010/40/EU and its corresponding Delegated Regulations, in particular:

- Commission Delegated Regulation (EU) No 2015/962 of 18 December 2014 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services
- Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to data and procedures for the provision, where possible, of road safety-related minimum universal traffic information free of charge to users

According to these Delegated Regulations, EU member states are required to consider and manage the quality of these services.

This framework supports EU member states in this context. Called a "Quality Package", it contains quality-related definitions and concepts, proposed and agreed by EU EIP partners for the use in Europe:

- Chapter 2 describes the Quality Criteria as basic parameters to describe quality of SRTI and RTTI services and the related data.
- Chapter 3 compiles the Quality Requirements as (minimum) quality levels to be reached by individual SRTI and RTTI services.
- Chapter 4 introduces potential methods for quality assurance and assessment, indicating their applicability for different uses and feasibility for studying different quality criteria.
- Chapter 5 contains conclusions on the scope and the proposed use of this quality package.

In addition to this "Quality Package", a supplemental document "Practical Guidelines" is provided. This document gives a short practical description on what to do in order to measure and to document the quality of traffic information, from a data supplier perspective.

2. Quality Criteria

2.1. Relation to Value Chain

The Quality Criteria, as introduced here, relate to traffic information within a specific part of the information process.

This information process can be illustrated by the Value Chain for SRTI (Safety-Related Traffic Information) and RTTI (Real-Time Traffic Information) services, as shown in Figure 1 below.

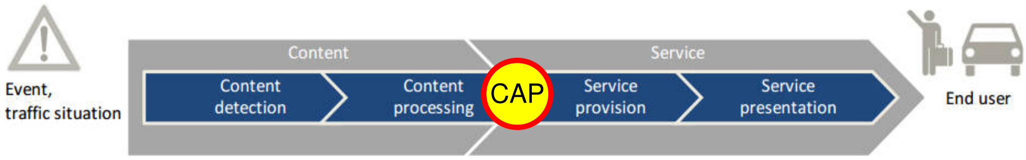


Figure 1. Value chain of SRTI and RTTI services

The Quality Criteria focus on the Content part of the Value Chain. The Content part, which is typically in the responsibility of a data supplier, covers processes between the detection of a real event or a traffic situation until the provision of related information in a Content Access Point (CAP). At a CAP, the traffic information is (typically) made available to many service providers via e.g. a data portal. This point can also be a National Access Point according to the Delegated Regulations of the EU ITS Directive.

On the other hand, aspects on the Service part of the Value Chain have been covered by the Traveller Information Services Association (TISA). TISA has published a Position paper on this (Ref 4), describing quality aspects as being important for the end users and to be met by service providers.

2.2. Quality Criteria definitions

A set of validated Quality Criteria and their corresponding definitions, as elaborated and validated in EU EIP, is shown below.

The shown set of Quality Criteria contains two criteria in the category 'Level of Service" (describing the provision of data) and ten criteria in the category 'Level of Quality" (describing the data as such).

Further, there is a differentiation between 'Event information' and 'Status/Entity information', because of the inherent differences in the nature of data related to these information types.

Table 1: Level-of-Service and Level-of-Quality Criteria for RTTI and SRTI

	Definition of Quality Criteria for RTTI and SRTI		Applicable for: Event Information	Applicable for: Status-Oriented Information
Level of Service	**Geographical coverage**	Percentage of the road network covered by the (content provision) service	X	X
Level of Service	**Availability**	Percentage of the time the (content provision) service is available	X	X
Level of Quality	**Timeliness (start)**	The time between the occurrence of an event and the acceptance of the event	X	-
Level of Quality	**Reporting period**	The time interval for refreshing / updating the status reports	-	X
Level of Quality	**Timeliness (update)**	The time between the end or (safety) relevant change of condition and the acceptance of this change	X	-
Level of Quality	**Timeliness (update)**	The average age of the sensor data used in the most recent reporting period	-	X
Level of Quality	**Latency (content side)**	The time between the acceptance of the event or its end or (safety) relevant change of condition and the moment the information is provided by the content access point	X	-
Level of Quality	**Latency (content side)**	The time between the calculation of the reporting data and the moment the information is provided by the content access point	-	X
Level of Quality	**Location accuracy**	The relative accuracy of the referenced location for the published event with respect to the actual location of the actual event	X	-
Level of Quality	**Classification correctness**	100% minus the percentage of the published events which are known to be not correct, concerning actual occurrence of this event type / class	X	-
Level of Quality	**Error Rate**	Percentage of published status reports which show excessive deviations of a reported quantity (e.g. speed or travel time) versus the actual value or are otherwise determined as erroneous	-	X
Level of Quality	**Event coverage**	Percentage of the events which are known to be correctly detected and published by type / class, time and location (i.e. detection rate)	X	-
Level of Quality	**Report coverage**	The percentage of reporting locations for which a status report is received in any given reporting period	-	X

2.3. Further explanation of time-related Quality Criteria

There are three time-related quality criteria ‘Timeliness’ (split up into ‘start’ and ‘update’), ‘Latency’ and ‘Reporting period’, which can be explained as follows (see also Figure 2 below):

- ‘Timeliness’ is the time span from the occurrence of an event until it is detected and accepted at the traffic centre.
- ‘Latency’ is the time span from the acceptance until the information (message) about the event is available at the CAP.
- ‘Reporting period’ is the so-called refresh rate of status reports.

There is a differentiation between the ‘acceptance’ and the ‘validation’ of an event. An event is considered accepted when it has been found trustworthy according to an organisation’s quality policy so that action will be taken for processing the event report and publishing the report at the Content Access Point (CAP). There may also be a validation process for a reported event, which can start/end either before or after the acceptance. Validation it is not used in the definitions presented here.

To guarantee comparable results it is strongly recommended that quality assessments involving ‘Latency (content side)’ document whether possible validation of the detection is included in the calculated latency values instead of timeliness. It is also recommended that latencies and timeliness are measured separately for the start, update and end of event.

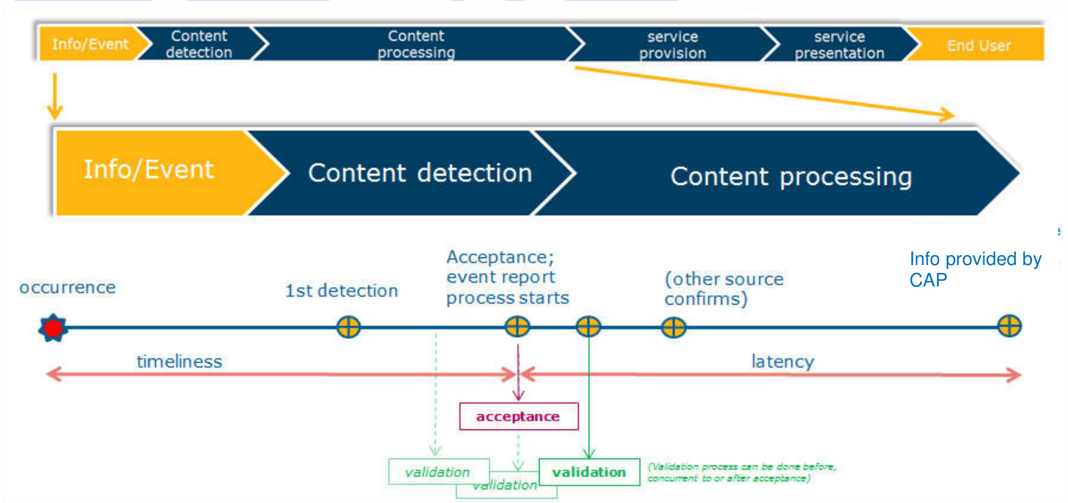


Figure 2. Value Chain with ‘Timeliness’ and ‘Latency’ indicated

Examples for the differentiation between ‘Timeliness (start)’ and ‘Latency (content side)’ with respect to necessary acceptance of information, are shown below in figures 3-6.

The following examples show how information is processed for a single event, including the differentiation between ‘timeliness’ and ‘latency’, see Figures 3-6. (Note that the shown lengths of ‘Timeliness’ and ‘Latency’ are for illustrative purposes only, and do not reflect actual durations.)

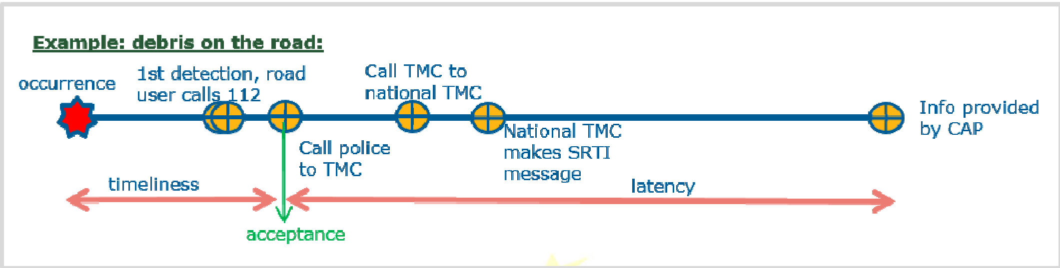


Figure 3. Debris on the road, road user calls 112.

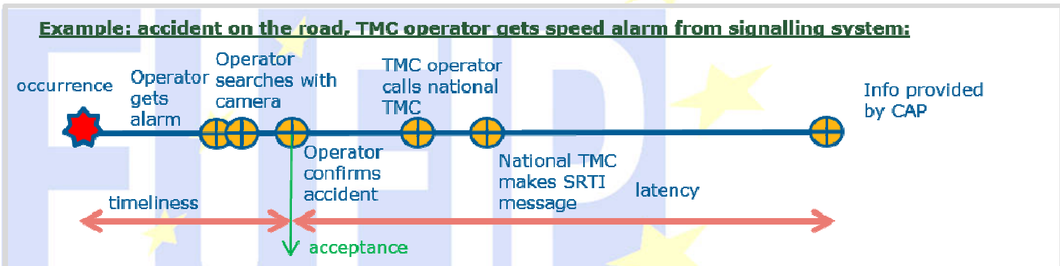


Figure 4.. Accident on the road, operator gets speed alarm from signalling system.

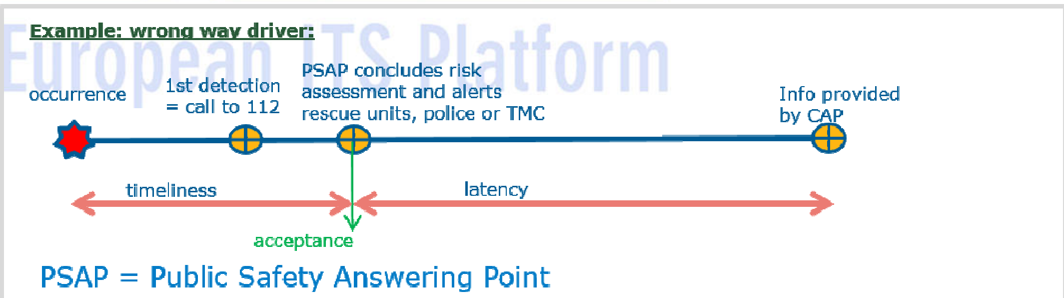


Figure 5. Wrong way driver, notification by an emergency call from a driver.

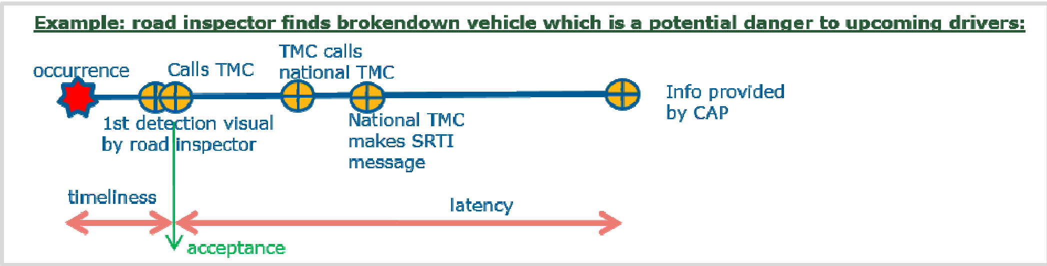


Figure 6. Road inspector finds broken down vehicle which is a potential danger to drivers.

2.4. Further explanation of 'Location accuracy'

'Location accuracy' indicates how correct the reported location is. The location accuracy shall be measured for the first message related to the event. Depending on how the information is transmitted, this Criterion may be applied to an area (e.g. an administrative area) or a link (e.g. a road section.)

The exact location is the actual, ground truth location indicating where the actual event was situated in the real-world. The reported location of the event in the SRTI or RTTI message may be:

- exactly the same as this exact location,
- less precise but still have an overlap with the exact location, or
- may not overlap with the exact location at all.

Still in all three cases, the reported location may be considered to be reported correctly as long as it falls within an "accuracy zone" specified in the criterion.

'Location accuracy' for links is specified either as "link between intersections" accuracy or as a metric criterion, depending on the Quality Requirements (see section 3).

For the "link between intersections" accuracy this means that the event must be located between the nearest downstream and upstream intersection location. However, there might be a "tolerance zone" allowing the location information to be "outside" of the downstream/upstream intersection stretch, as shown in Figure 7.

Such a "tolerance zone" may be useful when intersections are not consistently geo-referenced both in the detection and in the message. For example, a high-accuracy detection (coming e.g. from FCD) will not be consistently geo-referenced when using LCL nodes in the message. This could result in systematic inaccuracies, which could be handled with such a "tolerance zone".

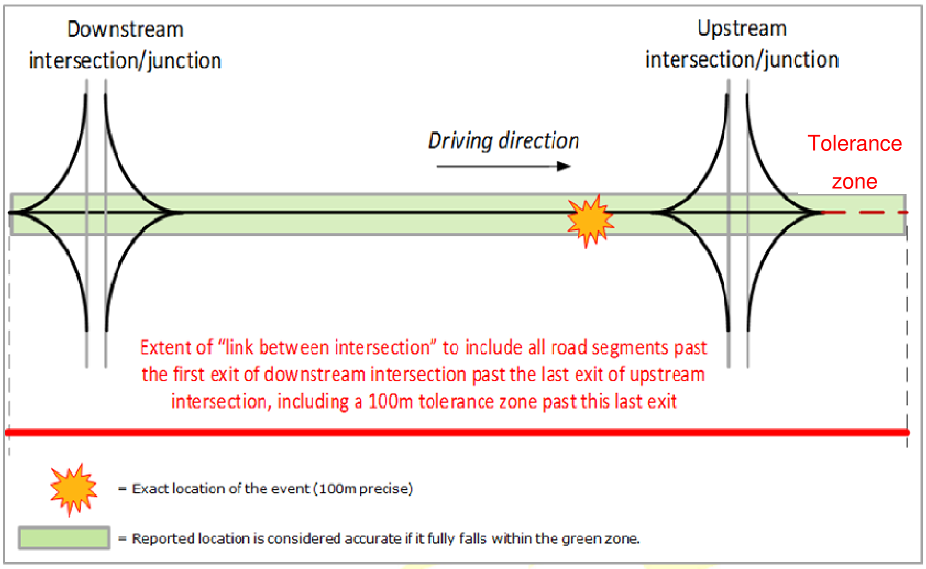


Figure 7. 'Location accuracy': Definition of "link between intersection accuracy" including a tolerance zone

Depending on the Quality Requirements, specific metric criteria should be used in addition to the "link between intersections" accuracy. Metric criteria, expressed in kilometres, show the geographic off-set of the actual vs. reported event location, regardless of the intersection reference.

Such a metric accuracy is useful for drivers, especially when embedded in navigation services. For example, an accident message may be communicated as "Caution, accident within the next 500 m !".

2.5. Further explanation of 'Classification correctness'

In case of classification correctness, inaccuracy is considered as an error only if it can have a consequence for the user behaviour. It is recommended to use time and place for matching events between data sets. Classification correctness shall be measured for the first message related to the event. Messages with label "preliminary" or similar shall not be considered as the first message. Messages which cannot be identified as correct or incorrect shall be excluded when calculating classification correctness.

3. Quality Requirements

This section compiles Quality Requirements for SRTI and RTTI services, as elaborated by EU EIP partners.

These requirements specify at which (minimum) level the Quality Criteria, as defined in section 2, should be realised by an individual SRTI and RTTI service to meet certain quality expectations. Such quality expectations are described in three levels:

** Basic*

*** Enhanced*

**** Advanced*

***** (Future technologies)*

The most important Quality Requirements are the minimum ones, denoted as the ** Basic* quality level. This level should be met by the services in all member states, because if the service would be provided at a lower level of service or quality, the user benefits would likely be negligible or even negative. In addition to the ** Basic* level, tentative quality requirement recommendations are also given for an *** Enhanced* and an **** Advanced* levels. Further, a quality level **** has been defined for Quality Requirements which will be related to near-future technologies, but it has not been specified with concrete values yet.

The Quality Requirements are expressed in a quantitative way where possible. They are understood as validated values, after being continuously discussed and evaluated during the recent work of EU EIP.

3.1. Level-of-Service Requirements

The Level-of-Service Requirements are defined equally for all SRTI/RTTI services, see Table 2.

Table 2: Level-of-Service Requirements for all SRTI/RTTI services

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★★ (Future technologies)
	Geographical coverage	**Best effort**	80%	95%	
	Availability	**95% (347 days/year)**	99% (361 days/year)	99,5% (363 days/year)	

3.2. Level-of-Quality Requirements for Real-Time Traffic Information (RTTI)

The Requirements are defined for different RTTI types, see the following tables:

- Table 3: Level-of-Quality Requirements for RTTI types: Real-time event information, Traffic conditions information (TCI), Weather information
- Table 4: Level-of-Quality Requirements for RTTI type: Travel Time information

Table 3: Level-of-Quality Requirements for RTTI types: Real-time event information, Traffic conditions information (TCI), Weather information

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★★ (Future technologies)
RTTI (real-time event, traffic conditions, weather)	Timeliness (start)[1]	**Best effort**	For 95 % of all events: Time between event occurrence and first detection: Best effort Acceptance after first detection < 10 min	For 95 % of all events: Detection & acceptance < 5 min after event occurrence	
	Timeliness (update)[1]	**Best effort**	Best effort	For 95 % of all events: Detection & acceptance < 10 min after event change/end	
	Latency (content side)[1]	**For 80% of all events: < 10 min**	For 80% of all events: < 5 min	For 95% of all events : < 5 min	
	Location accuracy - Area[2]	**For 95 % of all events**: **Correct administrative region**	For 95 % of all events: Correct geographic area; 10 km accuracy	For 95 % of all events: Correct geographic area; 5 km accuracy	
	Location accuracy - Road[3]	**For 95 % of all events**: **Correct link between Intersections** [4]	For 95 % of all events: Correct link between Intersections AND Off-set < 4 km [4]	For 95 % of all events: Correct link between Intersections AND Off-set < 2 km [4]	
	Classification correctness	**> 85%**	> 90%	> 95%	
	Event coverage	**Best effort**	Best effort	> 80% of all occurring events	

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★ ★ (Future technologies)
	[1] *For the first two levels (basic and enhanced), it is only necessary to measure timeliness and latency at the beginning and end of an event.* [2] *Not applicable for Traffic conditions information (TCI).* [3] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.* [4] *A tolerance zone of 500 meters to both directions from the geographical centre point of the intersection can be applied (e.g. to address differences in georeferencing systems). Ramps leading to rest areas only are not defined as intersections. The off-set requirement is not applied when only link-to-link information (i.e. without coordinates) is provided.*				

Table 4: Level-of-Quality Requirements for RTTI type: Travel Time information

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★ ★ (Future technologies)
RTTI (travel time)	Reporting period	**Best effort**	5 min	1 min	
	Timeliness (update) [1]	**Best effort**	For 95 % of all reports: < 5 min	For 95 % of all reports: < 2 min	
	Latency (content side)	**For 95 % of all reports**: **< 10 min**	For 95 % of all reports: < 5 min	For 95 % of all reports: < 2 min	
	Error rate	**Max. 10% of reports with deviation > 20 % OR being otherwise classified as erroneous[2]**	Max. 5% of reports with deviation > 20% OR being otherwise classified as erroneous[2]	Max. 5% of reports with deviation > 10% OR being otherwise classified as erroneous[2]	
	Report coverage	**Best effort**	> 90%	> 97%	
	[1] *'Timeliness (update)' may be left out as a criterion for Travel Time Information* [2] *The user might define own rules to determine if a report is classified as erroneous or not. These rules might come from e.g. sensor specifications or from internal policies.*				

3.3. Level-of-Quality Requirements for Safety-Related Traffic Information (SRTI)

The Requirements are defined for different SRTI types, see the following tables::

- Table 5: Level-of-Quality Requirements for SRTI type: Wrong Way Driver
- Table 6: Level-of-Quality Requirements for all SRTI types except Wrong Way Driver

Table 5: Level-of-Quality Requirements for SRTI type: Wrong Way Driver

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★★ (Future technologies)
SRTI (Wrong Way Driver)	Timeliness (start) [1]	**Best effort**	For 95 % of all events: Time between event occurrence and first detection: Best effort Acceptance after first detection < 5 min[1]	For 95 % of all events: Detection & acceptance < 2 min after event occurrence	
	Timeliness (update) [1]	**Best effort**	Best effort	For 95 % of all events: Detection & acceptance < 5 min after event change/end	
	Latency (content side) [1]	**For 80% of all events: < 5 min**	For 80% of all events: < 2 min	For 95% of all events: < 2 min	
	Location accuracy - Road	**For 95 % of all events**: **affected road segment not longer than 50 km or road number with destinations (city names) nearby**	For 95 % of all events: Correct link between intersections[2]	For 95 % of all events: Correct link between intersections[2]	
	Classification correctness	**> 75%**	> 85%	> 90%	
	Event coverage	**Best effort**	Best effort	> 80% of all occurring events	

[1] For the first two levels (basic and enhanced), it is only necessary to measure timeliness and latency at the beginning and end of an event.

[2] A tolerance zone of 500 meters to both directions from the geographical centre point of the intersection can be applied (e.g. to address differences in georeferencing systems). Ramps leading to rest areas only are not defined as intersections.

Table 6: Level-of-Quality Requirements for all SRTI types except Wrong Way Driver

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★★ (Future technologies)
SRTI (all SRTI events/ conditions except wrong way driver)	Timeliness (start)[1]	**Best effort**	For 95 % of all events: Time between event occurrence and first detection: Best effort Acceptance after first detection < 10 min[1]	For 95 % of all events: Detection & acceptance < 5 min after event occurrence	
	Timeliness (update)[1]	**Best effort**	Best effort	For 95 % of all events: Detection & acceptance < 10 min after event change/end	
	Latency (content side) [1]	**For 80% of all events: < 10 min**	For 80% of all events: < 5 min	For 95% of all events: < 5 min	
	Location accuracy - Area	**For 95 % of all events:** **Correct administrative region**	For 95 % of all events: Correct geographic area; 10 km accuracy	For 95 % of all events: Correct geographic area; 5 km accuracy	
	Location accuracy - Road	**For 95 % of all events:** **Correct link between intersections**	For 95 % of all events: Correct link between Intersections AND off-set < 4 km [3]	For 95 % of all events: Correct link between Intersections AND off-set < 2 km [3]	
	Classification correctness	**> 85%**	> 90%	> 95%	
	Event coverage	**Best effort**	Best effort	> 80% of all occurring events	

	Quality Criterion	★ (Basic)	★★ (Enhanced)	★★★ (Advanced)	★★★★ (Future technologies)
	[1] For the first two levels (basic and enhanced), it is only necessary to measure timeliness and latency at the beginning and end of an event.[2] For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria. *[3] A tolerance zone of 500 meters to both directions from the geographical centre point of the intersection can be applied (e.g. to address differences in georeferencing systems). Ramps leading to rest areas only are not defined as intersections. The off-set requirement is not applied when only link-to-link information (i.e. without coordinates) is provided.*				

4. Quality Assessment Methods

4.1. Descriptions of Quality Assessment Methods

4.1.1. Method 1: Continuous monitoring of equipment performance and availability

Method description

The method is intended for continuous monitoring of the functioning of the existing detector network. The aim is to get timely alert of the malfunctioning equipment in order to fix or replace them. The monitoring process may be automated or performed by a human user. The monitoring of equipment performance may include:

- verification of the availability of the data produced by the equipment
- check of consistency between the data values measured by the same equipment
- comparison of the measured data to other equipment in adjacent or to the same geographical area
- monitoring of error messages and alerts generated by the equipment.

The methods used to detect failures are more or less specific for the type of equipment. The detection of faults in inductive loops has been discussed in a literature review published in 2008 (Lu 2008), and the topic has been analysed further in a study report published in 2010 (Lu et al. 2010). The following checks for errors in loop data are provided as examples (Chen et al. 2003):

- The number of samples in a day with zero occupancy must be less than certain threshold
- The number of samples in a day with occupancy more than zero and flow equal to zero must be less than certain threshold
- The number of samples in a day that have occupancy greater than a certain value must be less than a certain threshold
- The entropy of occupancy samples must be greater than a certain threshold.

Loop detector faults and possible detection methods have been summarised by Lu et al. (2010, Table 4.1) (Table 7).

Table 7: Loop detector faults and fault detection methods (adapted from Lu et al. 2010).

Fault type	Possible detection methods
• No or insufficient data in a region of freeway section	• Traffic management centre (TMC) level aggregated data analysis to investigate a communication problem or a faulty loop controller. • Automatic communication error detection • Data quality checking for occupancy and flow

Fault type	Possible detection methods
• Synchronised error on a district, a freeway corridor or a section of freeway	• TMC level aggregated data analysis to investigate a communication problem or a power system failure • Automatic communication error detection • Data quality checking for occupancy and flow
• Inconsistent speed/occupancies for adjacent lanes persistently even in night hours	• Locate suspicious loops and controller cabinet for a field visit. • Data quality checking for occupancy and flow
• No data from some individual loops	• Possible causes: open circuit in loop, electrical power failure, missing parts, loop disconnected for some reason • Compare loop data with video data, locate suspicious loops and perform a field inspection • Data quality checking for occupancy and flow
• Insufficient data from certain loops	• Perform TMC level data analysis to locate suspicious loops and to exclude data communication link failure. • Automatic communication error detection • Data quality checking for occupancy and flow
• Wrong assignment in highway direction, between lanes, upstream and downstream or a dual loop	• Perform TMC level aggregated data analysis to locate suspicious loops and controller cabinet; perform a field inspection to check flow from loops in adjacent lanes. • Data quality checking for occupancy and flow
• Crosstalk	• The problem may be caused by interference between loop detector cards. • Perform a field inspection and compare neighbour loop signals. • Data quality checking for occupancy and flow
• Pulse flickering, chattering, misfiring	• Field inspections, offline data analysis for a long time period • Data quality checking for occupancy and flow
• Inconsistent data quality from time to time and from lane to lane without data link problems	• Field inspections, offline data analysis. • Data quality checking for occupancy and flow

Continuous checking of monitoring or detection equipment performance has also been applied to road weather stations. For example, the data quality checking algorithm used for weather data in the US Clarus initiative (U.S. DOT 2014) has been documented on a general level by Limber et al. (2010). A summary of the tests included in the algorithm is presented in Table 9.

Table 8: Tests included in Clarus data quality checking algorithm (summarised from Limber et al. 2010).

Test name	Description
Sensor range test	Values measured by a sensor are tested against the sensor hardware specifications or theoretical limits.
Climate range test	Values measured by a sensor are tested against predetermined climate range values.
Time step test	This test monitors the difference between the measured value and the previous measured value or the rate of change within a defined period of time.
Like instrument test	Sensor readings are compared to readings of similar sensors within the same road weather station.
Persistence test	The test detects sensor readings which stay constant for a long period of time. If consecutive sensor readings over a predetermined period are the same, the test does not pass.
Interquartile range (IQR) spatial test	The test checks the consistency of a sensor reading with neighbouring observation stations measuring the same variable. The IQR test does not pass when: $$\lvert Z_e - Z_0 \rvert > \max(M \times 0.7413 \times IQR, minToleranceBound)$$ where Z_e is the median of neighbours, Z_0 is the sensor reading under analysis, M is a constant multiplier value (3 for all data types except 2.5 for relative humidity), IQR is the interquartile range (the difference between 25th and 75th percentiles of neighbours and minToleranceBound is a fixed constant which determines the minimum acceptable difference between the data value under analysis and the estimate.
Barnes spatial test	The observation does not pass the Barnes spatial test when the observation differs more than a configured number of standard deviations from neighbouring observations. This can be expressed with equation: $$\lvert Z_e - Z_0 \rvert > sdMin \times \sigma$$ where Z_e is the weighted mean of neighbouring observations, Z_0 is the sensor reading under analysis, sdMin is the number of configured standard deviations and σ is the standard deviation of neighbouring observations. The weighted mean of neighbouring observations (Z_e) is calculated with equation: $$Z_e = \frac{\sum W(r_i) \times Z_i}{\sum W(r_i)}$$ where:

Test name	Description
	$$W(r_i) = e^{\frac{-\lvert Z_i - Z_0 \rvert^2}{2\times\left(\frac{r_i}{\sigma}\right)^2}}$$ and Z_i is the ith neighbouring observation.
Dewpoint temperature test	Once air temperature and relative humidity readings are obtained, dew point temperature is calculated. The estimated dew point temperature is then compared using the Barnes spatial test.
Sea level pressure test	The test is used to check whether atmospheric pressure readings are consistent with neighbouring observations. The measured pressure values are transformed to values present on sea level to allow comparison. The details of the test are documented in (Limber et al. 2010).
Precipitation estimation test	The procedure can be used to test whether the measured value of precipitation is within the minimum and maximum acceptable values. The procedure is described in appendix C of (Limber at al. 2010).

Data requirements

The method requires access to real-time data generated by the equipment under monitoring and the possible error messages or alerts. Some tests (Table 8) require also data which can be used for testing the data under analysis. In addition to the application layer, access to lower layers of data transmission may be needed for example when monitoring the status of the data link between roadside equipment and back-office system.

Applicability

Quality assurance

The method is used for quality assurance. It is used for following-up the monitoring systems deployed by the TCC.

Parts covered value chain

This method could be used in the different phases. It is mainly used in the content detection phase (Figure 7), but it could also be used in the content processing and service provision phase.

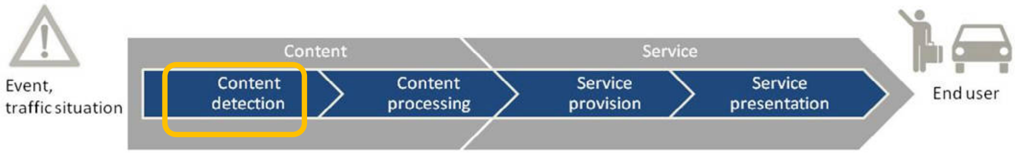


Figure 7. Parts of value chain covered by the method – Continuous monitoring of equipment performance and availability..

Type of service aspect / service equipment

This method could be used to monitor different types of equipment. It is not only focused on monitoring systems (loops, meteorological stations,…) but also other equipment involved in the ITS service provision chain (for example VMS).

Covered criteria

Criteria that can be assessed are: Availability, Classification correctness, Event coverage and Report coverage

<u>Usage</u>

Objective

This method could be used to determine the quality assessment of service, but it could be also used as an acceptance test (for example if the installed system has a minimum percentage of availability)

What stage of the process

This method is used during operation to assess if the system or the service responds to the expectations. It could be used as an acceptance test to assess the system.

What rate of use

It could be used in the three rates of use. Usually it is used on single or individual spots when some deviations in the monitoring data are detected.

Parts covered value chain

This method could be used in the different phases. It is mainly used in the content detection phase, but it could also be used in the content processing and service provision phase.

Type of service aspect / service equipment

This method could be used to monitor different types of equipment. It is not only focused on monitoring systems (loops, meteorological stations,…) but also other equipment involved in the ITS service provision chain (for example VMS).

Covered criteria

Criteria that can be assessed are: Availability, Classification correctness, Event coverage and Report coverage

Usage

Objective

This method is used for quality assessment of service.

What stage of the process

This method is used during operation to assess if the system or the service responds to the expectations. It could be used as an acceptance test to assess the system.

What rate of use

It could be used in the three rates of use. Usually is used on single or individual spots when the system is just installed or when some deviations in the monitoring data are detected.

Table 9: Usage – Continuous monitoring of equipment performance and availability.

Method 1: Continuous monitoring of equipment performance and availability							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	☐	✓	Single / spot check in case of problems	☐
Acceptance testing	✓	operation	✓	☐	☐	Periodic	monthly or more frequently
Feasibility / testing new procedure / algorithm	✓	problem	✓	☐	☐	Continuous use	✓
Internal quality control / monitoring	✓	diagnosis	✓	☐	☐	Needed special equipment	software running in the back office
						Needed special knowledge	specifications of equipment
						Expected cost	
Remarks:							

Table 10: Applicability – Continuous monitoring of equipment performance and availability.

Method 1: Continuous monitoring of equipment performance and availability								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	✓	Equipment	✓	Geographical coverage	☐	*Choose an item.*
Assessment	☐	Content processing	☐	e.g. monitoring of loop detectors or road weather stations		Availability	✓	objective
						Timeliness start	☐	
Event / status		Service provision	☐	Process	☐	Reporting period	☐	*Choose an item.*
				If yes, give description.		Timeliness update	✓	results do encompass criteria
						Latency	☐	
Event	✓	Service presentation	☐			Location accuracy	☐	
Status	✓					Reporting accuracy	✓	quantitative
Offline / online						Classification correctness	☐	
Offline	✓					Event coverage	✓	quantitative
Online	✓					Report coverage	☐	
Remarks								

4.1.2. Method 2: Manual verification of events or conditions

Method (short name) and short description

Manual verification of events or conditions based on current actual reality

Short description

The manual verification focuses on correctness of reported event occurrence or reported conditions. It is mainly used for verification of manually reported events or conditions.

Relevant questions are: Does an event occur (at the reported location)? Is the reported type and dimension of event or condition correct? Is the reported location of event or condition correct?

Used methods to check information against the actual reality depend on personal and technical equipment. If CCTV cameras do exist at the respective road section, these can be used for manual verification. Otherwise, this can be done by field inspection. In Germany, for instance, the road traffic police verifies reported safety-relevant events or conditions by road inspection in line with danger prevention.

Applicability

Quality assurance / assessment

The method is mostly used for quality assurance to correct or delete wrong messages about actual reported events by responsible operators.

Event / status

The method is only practicable for event information.

Offline / online

The described method is only practicable as an online process.

Parts covered value chain

The applicability of manual verification is shown in the following diagram (Figure 8).

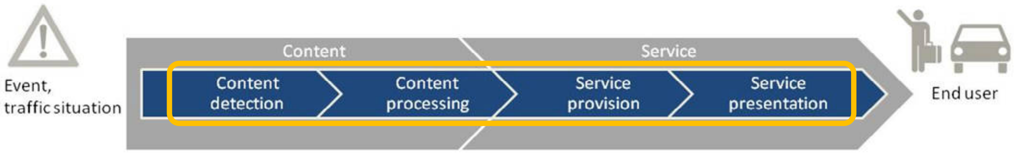


Figure 8. Parts of value chain covered by the method – Manual verification of events and conditions.

Manual verification can be used in all phases for quality assurance. It is mainly used to verify the quality in the content processing phase to avoid provision of wrong information based on manual detection.

Type of service aspect / service equipment

Manual verification can be based on CCTV cameras. Furthermore, a road inspection at the reported location is applicable.

Covered criteria

Criteria that can be assessed are: timeliness (focused on time for verification), location accuracy and classification correctness.

Results related to the criteria

The results of manual verification are:

- Qualitative (yes or no answering questions mentioned above)
- Objective (result is independent from the verifying human operator)
- Direct (positive or negative verification identifies correct or wrong messages)
- Encompass criteria (by decreasing of location accuracy the classification correctness increases).

<u>Usage</u>

Objective

Manual verification is mainly used for internal quality control, but other objectives can be addressed as well.

What stage of the process

Manual verification is used during operation to assess whether current reported events or conditions are correct or not. The method is based on the actual reality.

What rate of use

Manual verification is used as quality assurance method continuously depending on occurrence of relevant events. Events reported separately by many independent sources or events with low safety impact may cause lower importance of manual verification. Spot checks in case of problems are another use case of this method.

Manual verification of events is used by road operators and traffic reporting offices to ensure the correctness of manually detected events.

Experiences and actual use

References

Gerlach, J., Seipel, S., Leven, J.: Falschfahrten auf Autobahnen. Final report, 2012.

Table 11: Usage – Manual verification of events and conditions.

Method 2: Manual verification of events or conditions							
Usage							
Objective		Stage of the process	used	useful	not useful	Rate of use / needs for usage	
Assessment of service	✓	acceptance test	☐	☐	☐	Single / spot check in case of problems	✓
Acceptance testing	✓	operation	✓	✓	☐	Periodic	not periodic
Feasibility / testing new procedure / algorithm	✓	problem	✓	✓	☐	Continuous use	✓
Internal quality control / monitoring	✓	diagnosis	✓	✓	☐	Needed special equipment	Possibly CCTV cameras
						Needed special knowledge	basic
						Expected cost	?
Remarks:							

Table 12: Applicability – Manual verification of events and conditions.

Method 2: Manual verification of events or conditions								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	☐	Content detection	✓	Equipment	✓	Geographical coverage	☐	qualitative
Assessment	✓	Content processing	✓	*If yes, give description.*		Availability	☐	objective
						Timeliness start	✓	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	direct
				If yes, give description.		Timeliness update	✓	results do encompass criteria
						Latency	☐	
Event	✓	Service presentation	✓			Location accuracy	✓	
Status	☐					Reporting accuracy	☐	
Offline / online						Classification correctness	✓	
Offline	☐					Event coverage	☐	
Online	✓					Report coverage	☐	
Remarks								

4.1.3. Method 3: Reference testing of data collected

Method description

Reference testing of collected data includes practices that are used to verify that traffic condition, travel time or event information produced by a certain method is correct. The data or information under analysis is compared against a source known to be reliable (ground truth). The comparison is made for a limited period of time or limited amount of data in the context of an existing traffic information service. Many times this methodology is linked to purchasing information from a private company or piloting, implementation of new data collection or processing methodology or when any other changes are introduced to the service.

The choice of mathematical methods to be used for comparison depends on the characteristics of data to be evaluated (1), the data used as ground truth (2) and the objectives of the study (3). Suggestions for analysis approaches for continuous, multinomial or binomial data which is a function of time in one or two dimensional space can be found in Table 14.

Table 13: Suggested analysis approaches for continuous, multinomial or binomial data which is a function of time in one or two dimensional space.

Data types (as a function of time and location)		**Suggested analysis approach**
Data under analysis	Ground truth	
Multinomial or binomial	Multinomial or binomial	Create a mapping between the data types in the data set under analysis and the data set used as ground truth. Then use a time-space oriented reference testing method such as QKZ or QRTTI in analysis.
Multinomial or binomial	Continuous	Recode the ground truth data into categories that are used in the data set under analysis. Then use a time-space oriented reference testing method such as QKZ or QRTTI in the analysis.
Continuous	Multinomial or binomial	Recode the data under analysis into categories used in the ground truth data. Then use a time-space oriented reference testing method such as QKZ or QRTTI in the analysis.
Continuous	Continuous	• Analyse the distribution of error and absolute error (absolute error = |ground truth – data under analysis|) for example by visual plotting and calculating basic statistics such as mean, median, minimum, maximum and different percentiles (for example, 85th, 90th 95th and 99th). • Alternative approach: Recode both the data under analysis and the ground truth data into categories and process as multinomial data.

The analysis becomes slightly different, if the data set under analysis or the data set to be used as ground truth contains information expressed for distinct locations (points in one or two dimensional space) instead of a one or two dimensional area. The suggested analysis approaches for cases in which either the data set under analysis or the ground truth contains point based data are presented in Table 15.

Table 14: Suggested analysis approach for comparisons involving point based data.

Data types (point or area oriented data)				**Suggested analysis approach**
Data under analysis		Ground truth		
Data values	Defined range	Data values	Defined range	
binomial or multinomial	points	binomial or multinomial	points	Select the points covered by both data sets for analysis. Calculate detection rate and false alarm rate over time (like with time-space oriented reference test methods). The detection rate and false alarm rate may be calculated individually for different points or for a group or points in an aggregated manner.
binomial or multinomial	points	binomial or multinomial	area	Create a mapping between points in the data set under analysis and corresponding subsets of the area for which the ground truth data is defined. Calculate detection rate and false alarm rate over time (like with time-space oriented reference test methods).
binomial or multinomial	area	binomial or multinomial	points	Create a mapping between points in the data set used as ground truth and corresponding subsets of the area for which the data set under analysis is defined. Calculate detection and false alarm rate over time (like with time-space oriented reference test methods).
binomial or multinomial	area	binomial or multinomial	area	Consider time-space oriented reference test methods.
Note: for pre-processing of continuous data, see Table 14.				

For data types with continuous values, there are at least two possible approaches for reference testing of data. First, continuous data values can be assigned into categories and then handled as binary or multinomial data using time-space oriented reference test methods. Second, it is possible to analyse the difference between the ground truth and the values in the data set under analysis. When analysing the difference – the error of the measured data – it can be possible to use the data quality attributes already defined in ISO technical report ISO/TR 21707 and the QUANTIS project and statistical methods available in textbooks.

It may be necessary to calculate confidence intervals for variables such as probability of detection or false alarm rate to obtain statistically conclusive results. In case of event-oriented information such as warnings, it may be possible to assume that the probability of successful detection is equal for all of the events to be reported and that the probability of detection is not dependent on the outcomes of previous similar events. In this case, individual events can be considered to be Bernoulli trials with probability p for detection, and confidence intervals for the probability of detection can be estimated.

In case of event-oriented information such as warnings, it may be possible to assume that the probability of successful detection is equal for all of the events to be reported and that the probability of detection is not dependent on the outcomes of previous similar events. In this case, individual events can be considered to be Bernoulli trials with probability p for detection, and confidence intervals for the probability of detection can be estimated.

Reference testing of collected data is a method which can be applied to both services and data without restrictions related to position in the value chain (Figure 9).

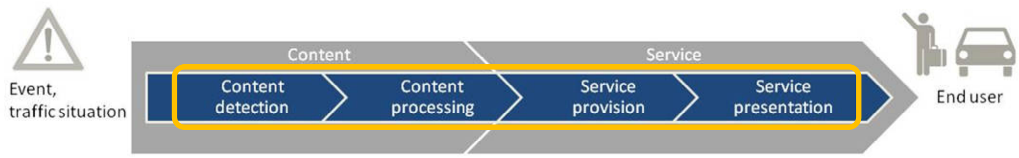


Figure 9. Parts of value chain covered by the method – Reference testing of data collected.

Data requirements

Reference testing of data requires a ground truth – data which can be considered to be correct with high probability. In addition to the ground truth, also other data sets may be used to support the conclusions of the analysis. The accuracy of analysis results is dependent on the quality of the data set used as a ground truth. If no data set considered suitable for ground truth is available, no reference testing can be performed. If the data set used as ground truth has substantial inaccuracies, the result of the whole analysis may be distorted.

In the following table, different alternatives for ground truth are mapped for each service type separately. The intention is to give a long list of possibilities to choose from, depending on the local available data. The choice for ground truth also depends on the data that is used in the information service itself. The ground truth should always be an independent set of data, so the mapping is only applicable, if the service is based on different data.

For RTTI events the current ground truth sources are separated as primary and secondary sources, where the primary sources are considered generally trustworthy. In many cases the use of secondary sources may require further verification or expert evaluation for suitability as ground truth addressing the local conditions. As a future perspective, potential sources are listed that may accompany the current sources to improve the coverage and quality of the ground truth data.

The SRTI ground truth sources are handled slightly differently, addressing the possibility to use sources for the evaluation of all occurring events or all reported events. This distinction points out to the fact that there are difficulties finding a suitable ground truth source for all occurred events for certain information services, because some of these events may last only a short time.

There is a variety of data sources that can be used as ground truth. It is advisable to search for intelligent combinations of data, e.g. listings of works by road maintenance operators combined with loop detector data about speed/flow reductions (that can verify the time and place the work took place at). In the near future, eCall will be a promising source for accurate incident data, at first for all reported events naturally. Vehicle probes used in cooperative ITS systems and automatic driving will eventually produce a vast dataset of event indicators that can be used as a ground truth for services that have been produced based on other types of data.

Table 15: Ground truth sources for RTTI information services.

	Current ground truth sources		Additional ground truth sources in the future
	Primary	Secondary	Potential
RTTI - Event information	• accident reports from police or insurance companies • other collections of occurred events	• automatic incident detection systems • crowd-sourcing services (e.g. Waze) • traffic reporting apps used by officials • news in media • social media (e.g. Twitter) • test drives (limited sample for assessment)	• eCall • cooperative systems, e.g. post-crash warning
RTTI - Traffic condition information	• test driving FVD • automatic number plate recognition systems • Bluetooth / WiFi monitoring systems • tolling systems	• commercial data (quality verified) • travel logs from various end-user devices	• any independent system with high amount of probe vehicles
RTTI - Travel time information	*same as above*	*same as above*	*same as above*
RTTI - Weather information	• independent field measurements / Road Weather Stations / FVD	• extended floating car data (e.g. windshield wiper activation data) • winter maintenance quality control data	• cooperative systems

Table 16: Ground truth sources for SRTI information services (1/2)

	Current ground truth sources		Additional ground truth sources in the future
	All Occurred Events	All Reported Events	Potential
SRTI - Temporary slippery road		• independent field measurements • road weather stations	• extended floating car data (e.g. friction sensors) • cooperative systems
SRTI - Animal/people/obstacles /debris on the road		• CCTV logs (limited sample) • 112 calls • news from media	• cooperative systems • automated driving • manual ecall
SRTI - Unprotected accident area	• accident reports from police or insurance companies • automatic incident detection systems	• emergency centre logs • accident reports from police or insurance companies • automatic incident detection systems • news from media • test drives / GPS dashboard cameras (limited sample for assessment)	• eCall (both automatic and manual) • cooperative systems • automated driving
SRTI - Short term road works	• listings of works ordered by the road authority • listings of works by road maintenance companies (actions such as grass cutting are not ordered separately each time) • test drives / GPS dashboard cameras (applies to limited network or sample)	• listings of works ordered by the road authority • listings of works by road maintenance companies • plausibility cross-check of the above against flow/speed reduction • traffic reporting apps used by officials	• cooperative systems • automated driving

Table 17: Ground truth sources for SRTI information services (2/2)

	Current ground truth sources		Additional ground truth sources in the future
	All Occurred Events	All Reported Events	Potential
SRTI - Reduced visibility	• road weather stations • CCTV logs on fully instrumented roads	• road weather stations • crowd-sourcing services, e.g. Waze • traffic reporting apps used by officials	• cooperative systems • automated driving
SRTI - Wrong-way driver	• CCTV logs on fully instrumented roads	• emergency centre logs • CCTV logs (limited sample) • news from media	• cooperative systems • automated driving • manual eCall
SRTI - Unmanaged blockage of a road	• tunnel and bridge management system logs • CCTV logs on fully instrumented roads	• tunnel and bridge management system logs • news from media • social media report with cross-check against implausible absence of traffic / probes	• cooperative systems • automated driving • manual eCall
SRTI – Exceptional weather conditions	• road weather stations • private weather stations	• road weather stations • news from media • crowd-sourcing services, e.g. Waze • extended floating car data (vehicle weather sensors)	• cooperative systems • automated driving

Expected results and scope of application

The method is applicable only in situations where it is possible to use data from an independent source as a ground truth. The data to be used as ground truth should be collected with means considered reliable and accurate. In other words, the ground truth data should be accurate and its possible weaknesses should be known at least on general level.

The application of the methods requires also that the ground truth data is representative of the conditions in which the system to be evaluated is supposed to be operating. For example, it may be preferable that the analysis and the ground truth data covers different traffic conditions in terms of traffic volumes or level of service, presence of incidents and different weather conditions.

Table 18: Usage – Reference testing of data collected.

Method 3: Reference testing of data collected							
Usage							
Objective		Stage of the process				**Rate of use / needs for usage**	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	✓	☐	Single / spot check in case of problems	✓
Acceptance testing	✓	operation	✓	☐	☐	Periodic	yearly
Feasibility / testing new procedure / algorithm	✓	problem	✓	☐	☐	Continuous use	☐
Internal quality control / monitoring	✓	diagnosis	☐	✓	☐	Needed special equipment	software for comparison of two data sets
						Needed special knowledge	data set to be used as ground truth
						Expected cost	€…..
Remarks							

Table 19: Applicability – Reference testing of data collected.

Method 3: Reference testing of data collected								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	✓	Equipment	✓	Geographical coverage	☐	*Choose an item.*
Assessment	✓	Content processing	✓	*If yes, give description.*		Availability	☐	objective
						Timeliness start	✓	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	*Choose an item.*
				If yes, give description.		Timeliness update	✓	results do encompass criteria
						Latency	✓	
Event	✓	Service presentation	✓			Location accuracy	✓	
Status	✓					Reporting accuracy	✓	
Offline / online						Classification correctness	✓	
Offline	✓					Event coverage	✓	
Online	☐					Report coverage	✓	
Remarks								

4.1.4. Method 4: Time-space oriented reference test methods

Method description

This group of methods consists of several methods, some well-established and widely used and some more experimental. With these methods, it is possible to compare the measured values in time and space – the data set under study – to the ground truth.

QKZ

QKZ ("Qualitätskennzahlen") (Bogenberger 2003c, Dance et al. 2008, Bogenberger and Weikl 2012) stands for Quality ratings.

The QKZ method focuses on correctness of reported event occurrence. Assessment of the traffic information messages' objective quality is based on reconstructing the actual traffic situation. The goal is to simulate as realistically as possible the actual situation experienced by the driver. It is advisable to gather independent traffic data of the data used to generate the traffic information, thus form the basis of the actual traffic situation's reproduction.

A brief description of QKZ has already been provided in EIP A3.2 final report. An example of the QKZ approach is presented in Figure 10. QKZ is one of the approaches used to assess the quality of traffic data and information. The approach has been originally described over ten years ago (Bogenberger 2003a). In summary, QKZ has two quality indicators: detection rate (QKZ1) and false alarm rate (QKZ2). The definitions of these indicators are available in a related conference paper (Bogenberger 2003b). Examples of their application have been provided by several authors (Li and Bertini 2010, Lux 2011, Bogenberger 2010). QKZ approach has been used by industry stakeholders such as Mediamobile (MediaMobile 2014). In the QKZ approach, the ground truth is typically established with data collected by roadside detectors, and the output of the service under evaluation such as a stream of congestion messages is compared against the ground truth.

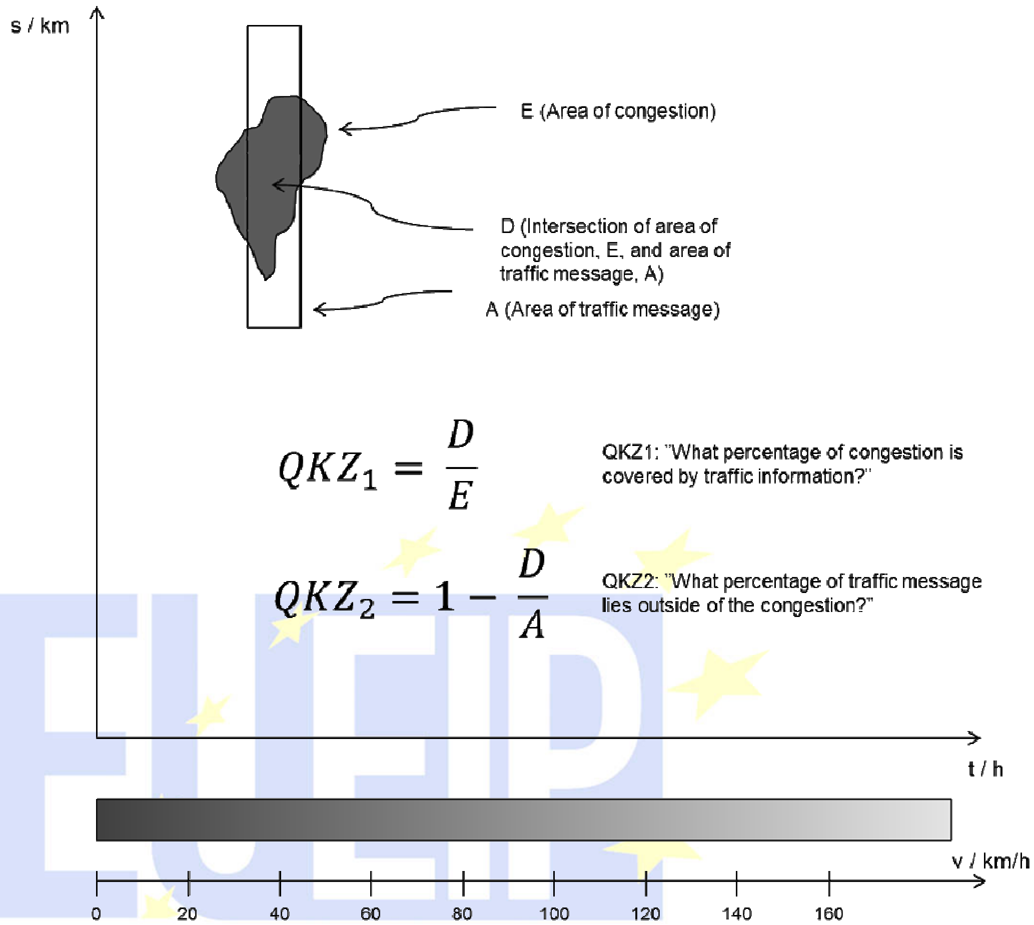


Figure 10. Example of the QKZ approach for assessment of the quality of traffic messages (adapted from Lux 2011).

Mathematical formulations of A, E and D can be found in an article published by Li and Bertini (2010):

$$A = \iint M(s,t)\, dsdt \qquad (1)$$

$$D = \iint C(s,t)\, dsdt \qquad (2)$$

$$E = \iint M(s,t)C(s,t)\, dsdt \qquad (3).$$

QKZ approach is typically applied to traffic congestion reports. It can be applied to categorical or binary data types.

QSRTI

QSRTI (quality assessment of event based, safety related, traffic information) is a method developed by the Dutch research institute TNO to address the characteristics of safety-related traffic information in the quality assessment.

The QSRTI method is described in TNO report 2015-10032 “QSRTI: a method for quality assessment of event based traffic information with a focus on Safety Related Traffic Information“ (Teun Hendriks, Carolien van der Vliet-Hameeteman 19 February 2015) and published in the 22nd ITS World Congress, Bordeaux, France, 5–9 October 2015: QSRTI: a method for quality assessment of event based (safety related) traffic information, Teun Hendriks, Carolien van der Vliet-Hameeteman, paper ITS-2050 .

The QSRTI method owes its heritage to the QKZ/QFCD methods developed by BMW, now widely used in the ITS industry. For the purpose of quality assessment for safety related traffic information, adaptations needed to be made to the QKZ method. These adaptations were necessary, firstly, to fit the quality criteria definitions by the EIP project and secondly, to be able to assess “minimum” type requirements, including accepted reporting tolerances and even some “best effort” values on the most basic level. (Hendriks & van der Vliet-Hameeteman 2015a and Hendriks & van der Vliet-Hameeteman 2015a)

The basic principle of the QSRTI method is a comparison of reference data (“ground truth”) and Content or Service Provider SRTI data (messages.) Specific adaptations have been made to include the EIP-defined tolerances in timeliness, latency, and location accuracy. Furthermore, to deal with proposed ‘best effort’ requirements for timeliness, and coverage at minimum levels, an additional quality index for processing latency (QSRTI-latency) is defined. Finally, another quality index for service availability (QSRTI-availability) has been created (Hendriks & van der Vliet-Hameeteman 2015a).

The two primary quality criteria for the SRTI method address the following two quality aspects (which are the same as for the QKZ method):

• Classification correctness: % wrong reports (‘false alarms’)
• Event coverage: % coverage right reported

As in QKZ, time delays in detection (timeliness) and processing (latency) will increase the classification correctness and decrease the event coverage. Correct reporting of an event is defined as satisfying the following three requirements:

1. Reported event type/description is representative for the same category (out of the 8 categories as defined by the delegated act) as the occurred event
2. The reported location falls within the actual location of the occurred event, enlarged by accepted tolerances

3. The time of reporting of the event falls within the time window of the actual occurrence of the event, enlarged with tolerances for timeliness and latency.

The mathematics to calculate the two quality indices QSRTI1 and QSRTI2 is similar to that of QKZ (QKZ1 and QKZ2), except that tolerances for the time and spatial dimensions of the event message are added to the actual location and time of the occurred event. These tolerances are shown in Figure 11.

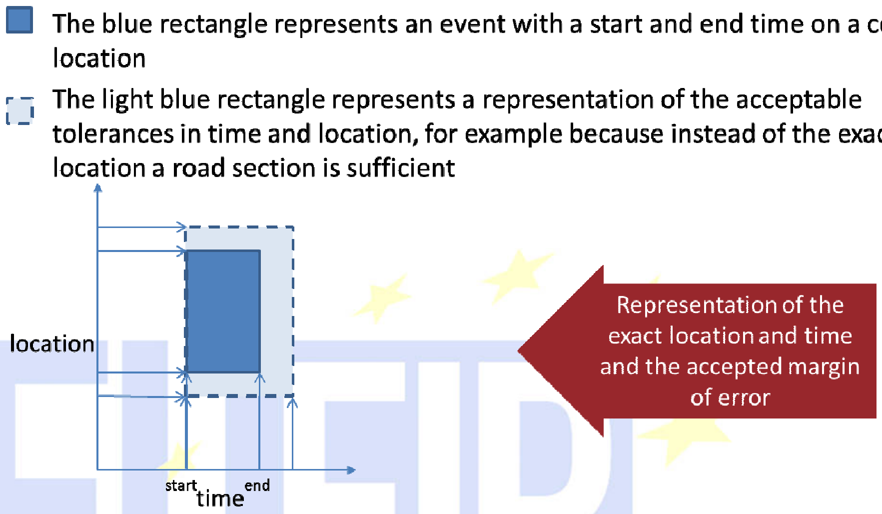


Figure 11. Definition of tolerance zone around actual occurred event in the QSRTI method.

A tolerance in location is added on both sides of the actual location. A tolerance in time (delay) is only added beyond the end-time of the actual event (and is the sum of the tolerance in timeliness and the tolerance in latency).

Figure 12 then shows the principle of assessment of an event message with respect the actual event and the tolerance zone. Assuming that the event category of the message and actual event are the same, then the 'false alarm' part of the message is only that area of the message which is completely outside the area of the event, and also completely outside the area of the tolerance zone. Assuming again that the event category of the message and actual event are the same, the message overlaps both with the actual event as with the tolerance zone are accepted as correct.

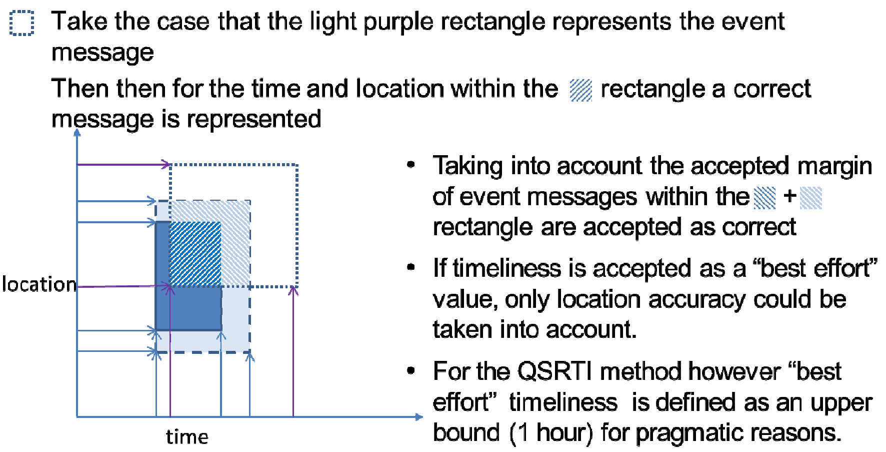


Figure 12. Assessment of an SRTI message versus an actual occurred event with tolerance zone.

Based on these tolerances, the quality criteria QSRTI1 (event coverage with tolerances) and QSRTI2 (classification correctness with tolerance) are computed. The exact formulas for computation of QSRTI1 and SQRTI2 can be found in literature (Hendriks & van der Vliet-Hameeteman 2015a and Hendriks & van der Vliet-Hameeteman 2015b).

<u>QRTTI – a modification of QKZ</u>

In the Finnish quality test carried out by Strafica Ltd for the Finnish Transport Agency, a modified QKZ method was used to assess the quality of incident information (Känsäkangas,2015).

In the test, a new incident detection algorithm was developed to detect abnormal incidents from the commercial link speed data provided by Here. The detection of the incident in the time and space dimensions from the Here data formed the ground truth to which the incident messages were compared. In the test, 50 % of the incidents reported were also detected by the method.

Event coverage and classification correctness criteria are calculated using a method modified from the QKZ for real time incident information. The challenge was to compare a point-dimensional incident message to two-dimensional abnormal congestion detected in the network.

In the measurement, the following assumptions are made:

- As the incident end time in FTA's reports is usually absent, a default value of 90 minutes from last update is used to define the end time. This is a justifiable assumption because the incident reports in Digitraffic service are available 90 minutes after which they automatically disperse if an update is not provided.

- The end of incident in the place-axis is defined in FTA's reports by the end of TMC link where the incident lies, even though most of the incident messages are reported as points. This was the marginal allowed in the space-axis.
- The end of incident in the place-axis in ground truth is defined by the end of exceptional congestion in Here speed data. FTA does not provide information about the area affected by the incident in its reports. Hence, it is assumed that instead of just reporting the location of the actual incident, FTA ought to report the end of queue due to incident.

Both ground truth and incident reports to be assessed are placed on a time-location space (figure 13). QRTTI1 is defined as a fraction of ground truth area where the incident message is valid

$$\mathrm{QRTTI}_1 = \frac{f(A \cap B)}{f(A)}$$

QRTTI2 is defined as a fraction of the message area not covering the ground truth

$$\mathrm{QRTTI}_2 = \frac{f(B) - f(A \cap B)}{f(B)}$$

where f is a function for area calculation.

The coverage C and classification correctness E of the entire study area in the entire study period is calculated as the arithmetic mean over all corresponding measures with the number of all traffic incidents (N).

$$C = \frac{\sum_{i=0}^{N} \mathrm{QRTTI}_1}{N}$$

$$E = \frac{\sum_{i=0}^{N} \mathrm{QRTTI}_2}{N}.$$

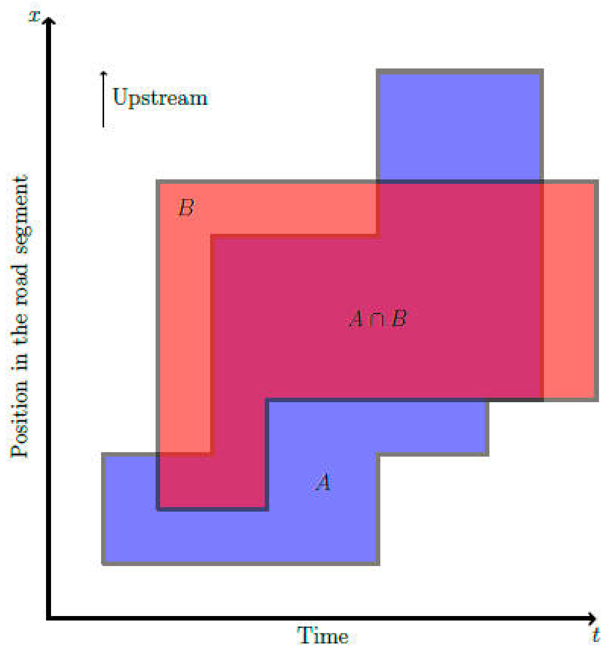


Figure 13. FTA incident information quality assessment setting using a modified QKZ-method. Red= message, blue = detected incident.

QFCD

QFCD is a microscopic model for measuring the quality of traffic Information. A description of QFCD has been published as a conference paper (Bogenberger and Hauschild 2009). QFCD has been used in the USA to verify Service Levels for Traffic Information Service providers, Dance et. al, 2008). The quality indicators used in QFCD are similar to QKZ: detection rate and false alarm rate (Q_fcd1 and Q_fcd2). Unlike in QKZ, no roadside data collection infrastructure is required QFCD approach.

The superimposition of recorded traffic states, received traffic messages and calculation of quality indices (Q_fcd1 and Q_fcd2) is illustrated in Figure 14. The status of traffic is reconstructed in the vehicle driving a test route using positioning data collected by an in-vehicle system. The traffic status measured by the test vehicle is presented with red rectangles and red lines in Figure 14. The traffic messages received such as congestion warnings received by the test vehicle are then compared against the measured traffic status. The traffic messages (in this case: congestion warnings) are presented with grey rectangles in Figure 14. The detection rate and false alarm rate can be calculated simply by comparing the areas A, D and E (Figure 14).

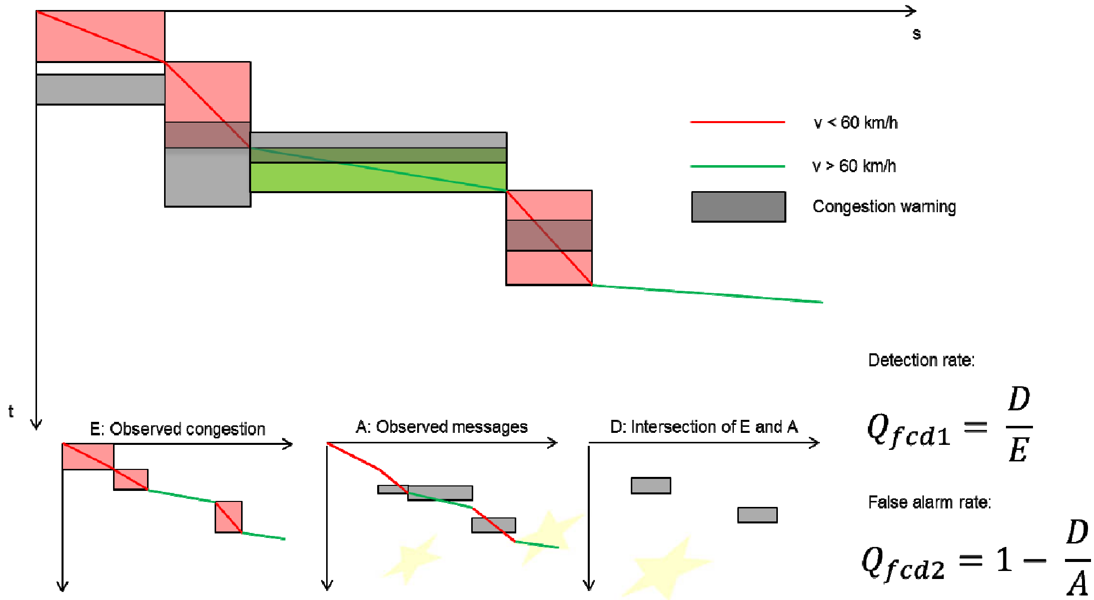


Figure 14. Example of QFCD approach (adapted from Bogenberger and Hauschild 2009).

When comparing the data to be evaluated with a data set considered to be correct with high probability, one has to make sure that the sample size is appropriate and that the sample is representative. First, one has to consider the effect of random variation when making a comparison between two data sets. Estimation of confidence intervals for the measured quality crtieria is recommended if this can be achieved with the data and tools available. Second, one has to take into account the representativeness of the analysed data set. For example, data collection equipment or analysis algorithms may perform differently in different traffic conditions.

For these presented methods, methodology exists for calculating detection rate and false alarm rate. However, methods for calculating confidence intervals for detection rate and false alarm rate are not publicly available. In practise, methods for estimating confidence intervals for quality crtieria such as those in QKZ should be selected on case-by-case basis depending on the characteristics of the data to be evaluated and the ground truth.

<u>Applicability</u>

Quality assurance / quality assessment

The method is mostly used for quality assurance, in that way it is used for testing and validation of new technical solutions, and the quality of existing information services eand service equipment.

Usage for quality assessment is nowadays visible in the determination of quality of data from other providers.

Parts covered in value chain

The applicability of QKZ and similar methods is shown in the figure below (Figure 15). The QKZ method can be used in the content detection phase in field testing of equipment. It is widely used to verify the quality after content provision phase or service provision phase.

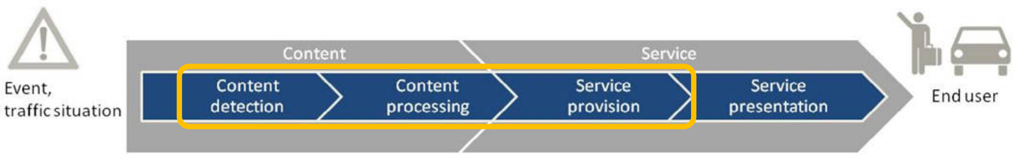


Figure 15. Parts of value chain covered by the methods

Type of service aspect / service equipment

Methods can be used for all types of equipment. The most important thing when using the methods is the availability of ground truth data. For each type of equipment or process event for which reference data is available the method can be used.

QKZ is widely used for testing the quality of RTTI services like congestion. QSRTI is to be tested in Netherlands to assess the quality of SRTI information services.

Covered criteria

Criteria that can be assessed are: Timeliness, Location accuracy, classification correctness and event coverage.

Results related to the criteria

<u>Usage</u>

Objective

Methods are used for quality assessment of services. Statistically relevant results can be obtained with this method.

What stage of process

Methods are used during operation to assess whether a service is delivered with good or bad quality.

What rate of use

Methods can be periodically used as spot check or yearly as assessment of SLA levels. For example, reference data of a set of highway segments for a certain period are used as ground truth.

Experiences and actual use

QKZ is in use by commercial service providers mainly to check quality of congestion reports. For example, Tom Tom has reported it uses QKZ.

Table 20: Usage – Time-space oriented reference test methods

Method 4: Time-space oriented reference test methods							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	☐	☐	Single / spot check in case of problems	☐
Acceptance testing	☐	operation	✓	✓	☐	Periodic	yearly
Feasibility / testing new procedure / algorithm	☐	problem	☐	☐	✓	Continuous use	☐
Internal quality control / monitoring	✓	diagnosis	☐	☐	✓	Needed special equipment	Equipment to collect ground truth
						Needed special knowledge	
						Expected cost	
Remarks:							

Table 21: Applicability – Time-space oriented reference test methods

Method 4: Time-space oriented reference test methods								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	☐	Content detection	✓	Equipment	☐	Geographical coverage	☐	quantitative
Assessment	✓	Content processing	☐	can indirectly indicate equipment faults		Availability	☐	objective
						Timeliness start	✓	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	indirect
				can indirectly indicate process faults		Timeliness update	✓	results do encompass criteria
						Latency	✓	
Event	✓	Service presentation	☐			Location accuracy	✓	
Status	☐					Reporting accuracy	☐	
Offline / online						Classification correctness	✓	
Offline	✓					Event coverage	✓	
Online	☐					Report coverage	☐	
Remarks								

4.1.5. Method 5: Monitoring of data completeness and latency

Method description

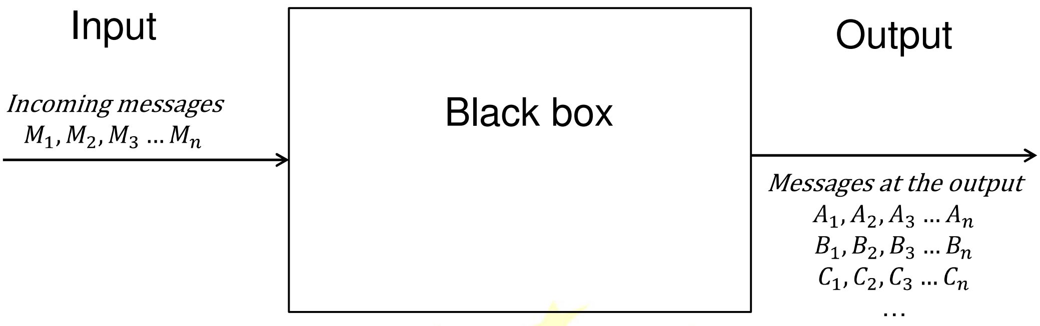
The objective of the automated monitoring of latency is to monitor the processing times of information in traffic information centre (TIC) or traffic management centre (TMC). It may also be implemented for other purposes. Automated monitoring of latency is typically implemented with software that automatically registers the time stamps of incoming/outgoing information related to a certain event within an organisation. This allows statistical analysis of the performance of the operator in the processing of the event and message provision.

The analysis can be divided into the following steps

1. Define the messages to be analysed and the part of service chain to be covered and treat it as a black box (Figure 16).
2. Describe the relation between messages at the input and messages at the output. The relation may be from one to one or from one to many. In the latter case, one has to define which of the messages at the output is chosen for analysis.

3. An automated system calculates the time difference between messages observed at the output and the corresponding messages at the input (Figure 16).

4. Aggregated key performance indicators are calculated and statistical analysis can be performed for the observed latencies of individual messages.



Latency l for message n which appears at the output: $l_n = t_{A_n} - t_{M_n}$

(note: message M at the input generates messages A, B and C; messages of type A selected for analysis)

Figure 16. Latency for message n which appears at the output of the system to be analysed.

Data requirements

The method requires an unambiguous relation between the messages at the input and output of the system to be analysed. The simplest case is the one to one relationship between input and output messages. Cases in which there is either one input message related to many output messages or many input messages related to one output message are more complex and require careful analysis. In these cases, there has to be objective criteria to determine which of the messages will be analysed.

The messages must also have reliable timestamp that is attached to a message when it enters or exits the system. This means that the software creating the input and output timestamps must be either using the same system clock or that the clocks used to create input and output timestamps must be either synchronised to each other or to a common external time reference.

4.1.5.1. *EXAMPLE – APPLICATION OF THE METHOD IN THE NETHERLANDS*

Method (short name) and short description

Monthly report

Short description

Every month the monthly report (factsheet) is generated which describes the availability, actuality and the overall score per data provider.

Availability: the amount of traffic data available from each data provider in the historical database with distinction in lane, location and minute. The percentage is calculated relatively to the total amount of expected traffic data:

- The total amount of expected traffic data is determined by the amount of minutes in one month, the amount of active locations during this month and the number of lanes per location. When locations are activated or deactivated during this month then they are not available during the inactive period.
- There is a distinction in available traffic data between error messages and usable traffic data.

Actuality: the time difference between the moment that a traffic data is available for publication in the central NDW system and the end of the measured minute.

- The actuality norm for the data providers is 55 seconds. Besides this period the central system needs 20 seconds to for staging the message. The actuality measured in the monthly report is 75 seconds. This is the content segment of the value chain.
- The actuality is only determined for the available traffic data. If only 1 part is delivered then the actuality is 100%.

Overall score: this gives an indication of the overall performance of each data provider for the delivery of actual traffic data. The overall score is determined by multiplying the availability and the actuality.

In the monthly report the result of each data provider for these indicators is presented and accommodated with an explanation.

Applicability

Quality assurance / assessment

Quality assurance / assessment

Event / status

At this moment, the monthly report only focusses on status traffic data. Event information is also provided by NDW but not yet monitored in the monthly report.

Offline / online

The method can easily be applied offline.

Parts covered value chain

In the Netherlands, the whole value chain of NDW (content segment) is examined for status services, like flow and travel times (Figure 17). The focus is on the completeness and not the content.

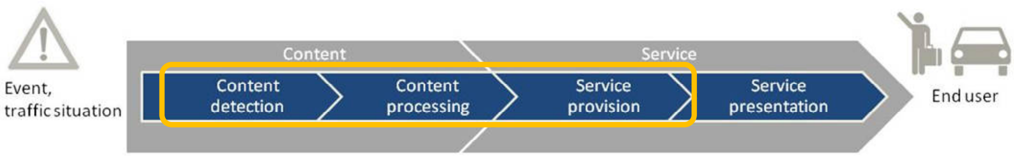


Figure 17. Parts of value chain covered by the method – Monitoring of data completeness and latency.

Type of service aspect / service equipment

The monthly report can be used for all types of equipment and processes. At this moment, the monthly report is only used for the delivery of status reports, but this is also possible for event reports. Starting point is the time stamp of every report.

Covered criteria

Criteria that can be assessed are: availability, timeliness start, reporting period, timeliness update, latency and classification correctness.

Results related to the criteria:

Quantitative
Objective
Direct
Results do encompass criteria.

<u>Usage</u>

Objective

The monthly report is used for assessment of service and internal quality control and monitoring.

What stage of the process

The monthly report is used to monitor the delivery of reports during operation. Besides, it can be used to identify problems and diagnose them.

What rate of use

The report is presented monthly to the data providers and the partners of NDW. The monitoring of the value chain is a continuous process.

Experiences and actual use

At this moment (Q4 2014), a process is started to improve the quality of the monthly report. In the current situation, the whole NDW value chain is monitored. Therefor is it difficult to improve the process because the value chain is divided into segments and for example, the delay can be in only one segment.

References

NDW, Factsheet kwaliteit verkeersgegevens juli 2014

Table 22: Usage – Monitoring of data completeness and latency.

Method 5: Monitoring of data completeness and latency							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	☐	☐	Single / spot check in case of problems	☐
Acceptance testing	☐	operation	✓	✓	☐	Periodic	monthly
Feasibility / testing new procedure / algorithm	☐	problem	✓	✓	☐	Continuous use	✓
Internal quality control / monitoring	✓	diagnosis	✓	✓	☐	Needed special equipment	
						Needed special knowledge	
						Expected cost	
Remarks: Monitoring is continuous, reporting is done in NL on a monthly basis							

Table 23: Applicability – Monitoring of data completeness and latency.

Method 5: Monitoring of data completeness and latency								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	✓	Equipment	✓	Geographical coverage	☐	quantitative
Assessment	✓	Content processing	✓	*If yes, give description.*		Availability	✓	objective
						Timeliness start	✓	
Event / status		Service provision	✓	Process	✓	Reporting period	✓	direct
				If yes, give description.		Timeliness update	✓	results do encompass criteria
						Latency	✓	
Event	✓	Service presentation	☐			Location accuracy	☐	
Status	✓					Reporting accuracy	☐	
Offline / online						Classification correctness	✓	
Offline	✓					Event coverage	☐	
Online	☐					Report coverage	☐	
Remark	In NL, atj NDW the whole value chain is examined (for status services, i.e. travel times) as an assessment.. method Completeness = Focused on quantity of reports/reporting location (not content in status report)							

4.1.6. Method 6: Regular sampling of message or data content completeness and correctness

Method description

Content samples of distributed traffic messages are once a month (manually) checked for correct message and data content.

Once a month a sample of around 20 % of distributed traffic messages of certain event types and around 20 % of certain operator processes are collected and checked by a person not having prepared the messages. The event types checked are objects on the road and accidents. Two processes are checked: Damage on road reports and traffic control (VMS) settings. Specified crtieria for the messages and reports are checked. The goal for the crtieria is that 90 % or more should be correct. If not, special emphasis is put on improvements for the next month. Other types of messages and processes as well as crtieria could be checked in the same way.

Applicability

Quality assurance/assessment

The method is for Quality assurance.

Event/status

The method is used for events and operator actions. It could also be used for status information.

Offline/online

The method is an offline method.

Parts covered in value chain

The operator actions, i.e. Content processing is covered (Figure 18).

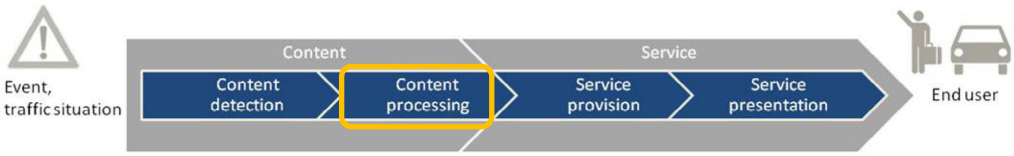


Figure 18. Parts of value chain covered by the method – Regular sampling of message or data content completeness and correctness.

Results related to the criteria

The results related to the criteria are:

Partly quantitative, partly qualitative
Objective
Direct
Results do not encompass criteria.

<u>Usage</u>

Objective

The objective is internal quality control / monitoring.

What stage of the process

The operational actions of the operators are checked.

What rate of use

The checking is done once a month. The frequency of testing could be different. The expected costs are not very high, since the checking is done by the TIC/TMC staff.

Experiences and actual use

The method has been in use since 2012 and has had a positive effect on the completeness and correctness of the operator actions.

Table 24: Usage – Regular sampling of message or data content completeness and correctness.

Method 6: Regular sampling of message or data content completeness and correctness							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	☐	acceptance test	☐	☐	✓	Single / spot check in case of problems	☐
Acceptance testing	☐	operation	✓	✓	☐	Periodic	monthly
Feasibility / testing new procedure / algorithm	☐	problem	☐	☐	✓	Continuous use	☐
Internal quality control / monitoring	✓	diagnosis	☐	☐	✓	Needed special equipment	
						Needed special knowledge	
						Expected cost	
Remarks:							

Table 25: Applicability – Regular sampling of message or data content completeness and correctness.

Method 6: Regular sampling of message or data content completeness and correctness								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	☐	Equipment	☐	Geographical coverage	☐	qualitative
Assessment	☐	Content processing	✓	*If yes, give description.*		Availability	☐	objective
						Timeliness start	✓	
Event / status		Service provision	☐	Process	✓	Reporting period	☐	direct
				If yes, give description.		Timeliness update	✓	results do not encompass criteria
						Latency	✓	
Event	✓	Service presentation	☐			Location accuracy	✓	
Status	☐					Reporting accuracy	✓	
Offline / online						Classification correctness	✓	
Offline		✓				Event coverage	✓	
Online		☐				Report coverage	☐	
Remark								

4.1.7. Method 7: Verification and calibration of traffic or weather conditions prognosis

Method description

The method allows constant verification of the prognosis regarding traffic conditions/travel time or road weather. The prognosis is systematically compared to the measured condition at the time in question, and the algorithm is calibrated accordingly. The methods used for weather and traffic forecasts are partly the same.

Overview of methods which can be used for weather forecast validation can be found in (Wilks 2005). Methods for forecast verification are provided for both for forecasts involving categorical and continuous variables. In case of a forecast with a non-probabilistic categorical variable, contingency tables and related measures can be used to illustrate and evaluate the quality of the forecast (Tables 28 and 29).

Table 26: Example of contingency table.

	Observed		
Forecast	Yes	No	
Yes	N_{11}	N_{12}	$N_{11}+N_{12}$
No	N_{21}	N_{22}	$N_{21}+N_{22}$
	$N_{11}+N_{21}$	$N_{12}+N_{22}$	

Table 27: Selected forecast quality metrics related to 2 x 2 contingency table (Wilks 2005).

Proportion correct	$\frac{N_{11}+N_{22}}{N_{11}+N_{12}+N_{21}+N_{22}}$
Threat score Critical success index	$\frac{N_{11}}{N_{11}+N_{12}+N_{21}}$
Odds ratio	$\frac{N_{11}\times N_{22}}{N_{12}\times N_{21}}$
Bias	$\frac{N_{11}+N_{12}}{N_{11}+N_{21}}$
False alarm ratio	$\frac{N_{12}}{N_{11}+N_{12}}$
Hit rate	$\frac{N_{11}}{N_{11}+N_{21}}$
False alarm rate	$\frac{N_{12}}{N_{12}+N_{22}}$
Heidke skill score	$\frac{\frac{N_{11}+N_{22}}{n}-\frac{(N_{11}+N_{12})(N_{11}+N_{21})+(N_{12}+N_{22})(N_{21}+N_{22})}{n^2}}{1-\frac{(N_{11}+N_{12})(N_{11}+N_{21})+(N_{12}+N_{22})(N_{21}+N_{22})}{n^2}}$

The conversion of probabilistic forecasts to non-probabilistic ones and extension of the contingency table approach to more variables and dimensions has been explained in (Wilks 2005).

The goodness of forecasts for a continuous variable can be illustrated with conditional quantile plots (Wilks 2005). An example of quantile plot is shown in Figure 19.

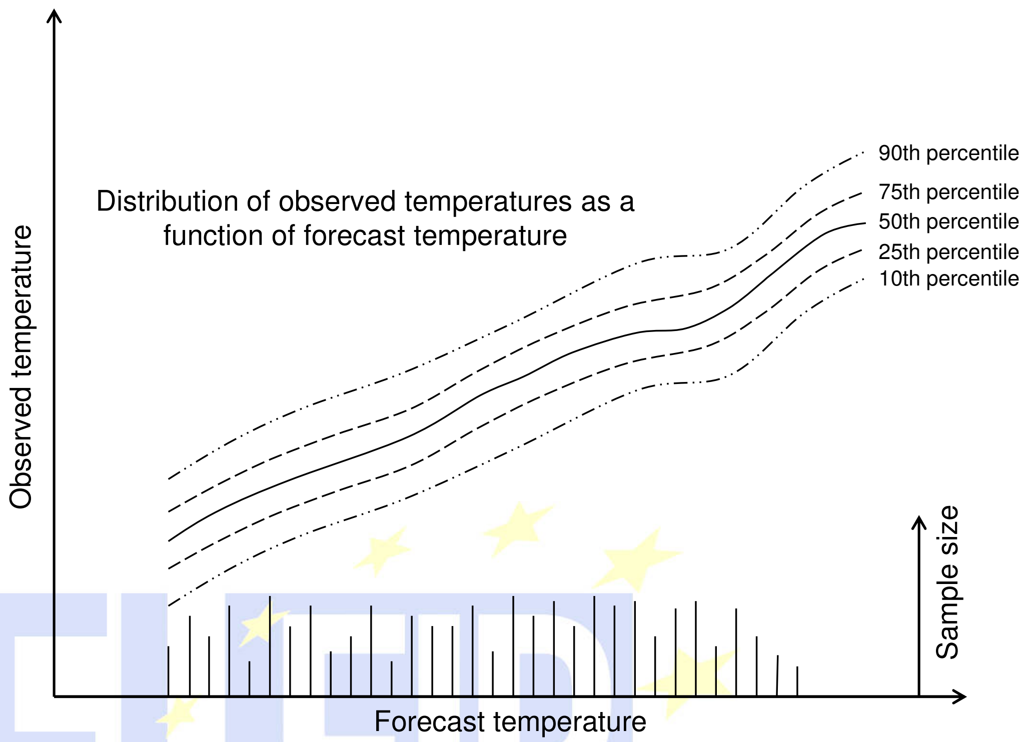


Figure 19. Example of quantile plot which shows the distribution of observed temperature as a function of forecast temperature (adapted from Wilks 2005).

The accuracy of a forecast for a continuous variable can be quantified with accuracy measures like mean absolute error or mean squared error (equations 3 and 4) (Wilks 2005).

Mean absolute error $MAE = \frac{1}{n}\sum_{k=1}^{n}|y_k - o_k|$ (3)

Mean squared error $MSE = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2$ (4)

where y_k and o_k are the pairs of forecast and observed values.

In addition, skill scores can be calculated for the forecast. Skill scores measure the goodness of the forecast in comparison to a reference forecast which can based on climatological data or historical values.

The verification of traffic status forecast can be carried out either using methods similar to the ones applied in meteorology or by using the QKZ or QFCD approaches.

Data requirements

The method requires the availability of both the forecast and the corresponding observed data. The sample size should be large enough to allow statistically robust conclusions. The method is applicable to service content but not data (Figure 20).

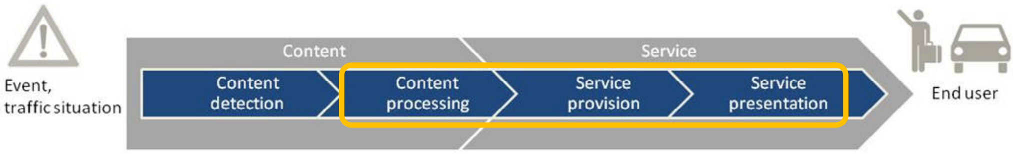


Figure 20. Parts of value chain covered by the method – Verification and calibration of traffic or weather conditions prognosis.

Table 28: Usage – Verification and calibration of traffic or weather conditions prognosis.

Method 7: Verification and calibration of traffic or weather conditions prognosis							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	✓	☐	Single / spot check in case of problems	✓
Acceptance testing	✓	operation	✓	✓	☐	Periodic	monthly
Feasibility / testing new procedure / algorithm	✓	problem	☐	✓	☐	Continuous use	☐
Internal quality control / monitoring	✓	diagnosis	☐	✓	☐	Needed special equipment	suitable software and data sets
						Needed special knowledge	knowledge on forecast valiation
						Expected cost	
Remarks:							

Table 29: Applicability – Verification and calibration of traffic or weather conditions prognosis.

Method 7: Verification and calibration of traffic / weather conditions prognosis								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	☐	Equipment	☐	Geographical coverage	☐	
Assessment	✓	Content processing	✓	*If yes, give description.*		Availability	☐	
						Timeliness start	☐	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	
				weather or traffic status forecast (e.g. travel time)		Timeliness update	☐	
						Latency	☐	
Event	✓	Service presentation	✓			Location accuracy	☐	
Status	✓					Reporting accuracy	✓	quantitative
Offline / online						Classification correctness	✓	quantitative
Offline	✓					Event coverage	✓	quantitative
Online	☐					Report coverage	✓	quantitative
Remark								

4.1.8. Method 8: Surveys of perceived quality by users

Method description

The aim of a user survey is to measure how the end users experience/perceive the traffic information services. Data collection may be performed periodically (e.g. once a year).

The degree of satisfaction, the degree of relevance, the user needs and the perceived quality are covered by the questionnaire. Many other questions are also asked. A web panel of a sufficient number of active car users are asked to participate in a survey (e.g. in Denmark, 1000 users - each driving more than 8000 km a year)..The services which can be covered by the method are: web sites, mobile applications, RDS-TMC, and teletext.

Applicability

Quality assurance/assessment

The method is for Quality assessment.

Event/status

The method covers the services, which means both event and status information.

Offline/online

The method is an offline method.

Parts covered in value chain

All parts of the value chain are covered (Figure 21).

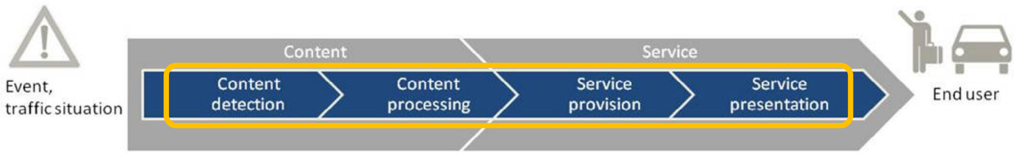


Figure 21. Parts of value chain covered by the method – Surveys of perceived quality by users.

Type of service aspect / service equipment

User surveys can be used for all types of services.

Covered criteria

In principle, all criteria are covered by the quality perceived by the end users: Availability, timeliness and latency (in total), location accuracy, classification correctness and event coverage. However, the users often cannot distinguish between the system down (availability), a missed event (event coverage), a wrong location and a long timeliness or latency, if they observe an event on the road, for which they have not received a traffic message.

Results related to the criteria

The results related to the criteria are:

- Quantitative
- Subjective
- Indirect
- Results do not encompass criteria.

Usage

Objective

The objective is assessment of the services.

What stage of the process

The operation of the services is judged.

What rate of use

The user surveys are performed once a year. The frequency of the surveys could be different. The surveys are done by commercial survey agencies. The costs per survey range from 10000 € to 40000 €.

Experiences and actual use

The mentioned surveys were used for the first time in 2012. In 2013, some improvements were observed. The 2014 results are not yet available.

<u>References</u>

Slides shown at and distributed after the EIP meeting on 21–22 November 2013 in Brussels. Internal documents in Danish.

Table 30: Usage – Surveys of perceived quality by users.

Method 8: Surveys of perceived quality by users							
Usage							
Objective		Stage of the process	used	useful	not useful	Rate of use / needs for usage	
Assessment of service	✓	acceptance test	☐	☐	☐	Single / spot check in case of problems	☐
Acceptance testing	☐	operation	✓	☐	☐	Periodic	yearly
Feasibility / testing new procedure / algorithm	☐	problem	☐	☐	☐	Continuous use	☐
Internal quality control / monitoring	✓	diagnosis	☐	☐	☐	Needed special equipment	
						Needed special knowledge	
						Expected cost	
Remarks:							

Table 31: Applicability – Surveys of perceived quality by users.

Method 8: Surveys of perceived quality by users								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	☐	Content detection	✓	Equipment	✓	Geographical coverage	☐	qualitative
Assessment	✓	Content processing	✓	*If yes, give description.*		Availability	✓	subjective
						Timeliness start	✓	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	indirect
				If yes, give description.		Timeliness update	✓	results do not encompass criteria
						Latency	✓	
Event	✓	Service presentation	✓			Location accuracy	✓	
Status	✓					Reporting accuracy	☐	
Offline / online						Classification correctness	✓	
Offline	✓					Event coverage	✓	
Online	☐					Report coverage	☐	
Remarks								

4.1.9. Method 9: Collection of direct user feedback

Method description

Collection of direct user feedback means using different channels established by the service provider to collect feedback from the users regarding the quality of the service in question.

In quality assessment, collection of direct user feedback is a relatively easy way to get information of how the actual users of the service experience the service quality. The feedback can be collected via webpage, where the feedback can be classified by the user and directed to the responsible parties. The feedback can also be collected by telephone, which requires a bit more resources for registering the feedback. User feedback is a very important method considering consumer information services (end user services), but can also be applied to b2b-type of services such as Content Access Point.

Applicability

Quality assurance / assessment

The method can be used as a part of the quality assurance process. If no other means of quality assurance are possible, at least this method should be used. The method provides only qualitative type of information which limits its use in systematical quality assessment.

Event / status

Method can be used for both types.

Offline / online

Method can be used in online analysis.

Parts covered in value chain

The coverage of the method in the ITS value chain is described in Figure 22.

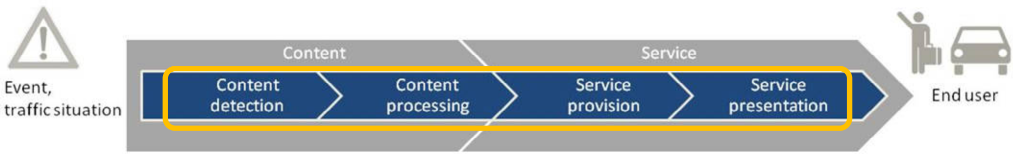


Figure 22. Parts of value chain covered by the method – Collection of direct user feedback.

Type of service aspect / equipment

The method is as its best in the evaluation of service as a whole from end users' perspective. The feedback, and the encountered problems, can however be traced to different phases in the value chain.

It should be kept in mind that the method is important also for the quality assurance of services such as Content Access Point, even though the users are companies and organisations using the interface.

Covered criteria

The method can be used to collect information about availability, location accuracy, reporting accuracy and perhaps classification correctness.

The feedback could also concern the delay in the value chain, but timeliness cannot be separated from latency. In addition, indications regarding poor event coverage can be achieved with this method.

Results related to the criteria

The collected feedback is qualitative in nature. Hence, if a lot of feedback is received concerning, for example, wrong travel times or wrong locations for incidents, that is a signal to make deeper quantitative analysis of the quality in the value chain. Therefore, the method does not directly measure the quality in terms of the defined attributes, but it collects (indirect) indications about the quality.

Feedback is always subjective.

<u>Usage</u>

Objective

Collection of direct user feedback is used for quality assessment of a running service. Because the method provides only qualitative information about the quality, it is not recommended as an only assessment method.

The method could also be used in the acceptance testing phase with a limited test group.

What stage of the process

The method is used in the operational phase of the service.

The method is also beneficial in the test phase of a new service. In this case, a test group of users is formed and their opinions are collected with the help of different channels.

What rate of use

The method is used continuously. Method does not require special expertise, but there are some minor costs related to the handling and analysis of the received feedback.

Experiences and actual use

Most services targeted for the end users are already using this assessment method. Possibility to leave user feedback is also part of a good customer experience, so the method is not only used as a means of quality assessment.

In a good practice of this method, the service provider explains the user for what purpose the feedback is collected, how it is processed, what the processing time of the feedback is, and whether the user will be provided with an answer from the service provider. In best practices, the user actually receives a personal answer including an explanation to what actions the feedback has/will lead to.

Hence, the use of the method requires person(s) who takes responsibility of the management of the feedback. The volume of users and the resulting amount of feedback defines the needed resources for this method.

A good practise to decrease the amount of unnecessary feedback is to set up a Questions&Answers page, where the most common feedback and the related answers are already addressed. Explanation about how the information is collected and what factors affect the quality of the information may be a good way to reduce the amount of (unnecessary) feedback.

The use of telephone in the collection is not recommended in case of end user services, because that would tie staff and requires more resources in the registration (typing in) of the feedback.

In case of b2b type of services such as Content Access Point, the amount of users is usually quite limited, and the feedback can also be collected via telephone as part of the customer service offered to the users of the service.

Table 32: Usage – Collection of direct user feedback.

Method 9: Collection of direct user feedback							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	✓	☐	Single / spot check in case of problems	☐
Acceptance testing	☐	operation	☐	✓	☐	Periodic	
Feasibility / testing new procedure / algorithm	☐	problem	☐	☐	☐	Continuous use	✓
Internal quality control / monitoring	☐	diagnosis	☐	☐	☐	Needed special equipment	
						Needed special knowledge	
						Expected cost	minor
Remarks:							

Table 33: Applicability – Collection of direct user feedback.

Method 9: Collection of direct user feedback								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	✓	Equipment	✓	Geographical coverage	☐	qualitative
Assessment	☐	Content processing	✓	can indicate equipment faults		Availability	✓	subjective
						Timeliness start	☐	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	direct
				can indicate process faults		Timeliness update	☐	results do encompass criteria
						Latency	☐	
Event	✓	Service presentation	✓			Location accuracy	✓	
Status	✓					Reporting accuracy	✓	
Offline / online						Classification correctness	✓	
Offline	☐					Event coverage	☐	
Online	✓					Report coverage	☐	
Remarks	Provides qualitative indication of the quality criteria							

4.1.10. Method 10: Monitoring of service use statistics

Short description

Monitor amount of service use to assess effect of service content and quality by using counters of internet page visits, smartphone application downloads and use etc.

The method provides only indirect information of service quality, but it is important as the main purpose of service quality is to provide benefit to the user of the service. The users will only use a service, if it provides such benefit, and thereby service use statistics are essential for the service providers.

Applicability

Quality assurance / assessment

Applies to both.

Event / status

Applies to both.

Offline / online

Primarily offline

Parts covered value chain

The method covers service provision and presentation, but indirectly all parts of the value chain (Figure 23).

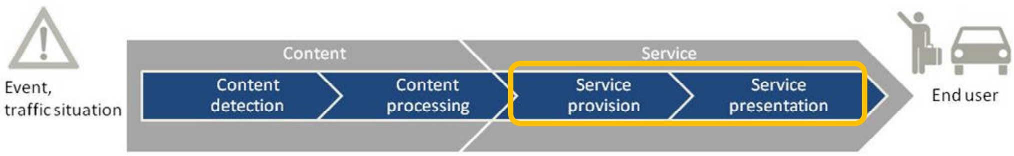


Figure 23. Parts of value chain covered by the method – Monitoring of service use statistics.

Type of service aspect / service equipment

Only the service as a whole, no possibility to differentiate between service aspects.

Covered criteria

Basically all as they all affect the whole service quality reflected in willingness to use service – it is very hard to establish link to availability, latency and timeliness criteria.

Results related to the criteria

Quantitative / qualitative: qualitative for the criteria, but quantitative as such

Objective / subjective: objective

Direct / indirect: indirect

Results only encompass criteria indirectly

Usage

Objective

The objective is to assess the benefit and thereby quality to the users by monitoring how often and by how many the service is being used.

What stage of the process

Service operation, useful in both normal and abnormal situations of service operation.

What rate of use

Applicable to single/spot check, periodic, and continuous use

No special equipment is needed except for software solutions to provide usage counters at internet sites and pages, as well as willingness of application and app store providers to provide information of application download and use statistics.

The costs are expected to be very low.

Experiences and actual use

The practice is widely used globally.

Table 34: Usage – Monitoring of service use statistics

Method 10: Monitoring of service use statistics							
Usage							
Objective		Stage of the process				Rate of use / needs for usage	
			used	useful	not useful		
Assessment of service	✓	acceptance test	☐	☐	✓	Single / spot check in case of problems	✓
Acceptance testing	☐	operation	✓	✓	☐	Periodic	✓
Feasibility / testing new procedure / algorithm	☐	problem	☐	✓	☐	Continuous use	✓
Internal quality control / monitoring	☐	diagnosis	☐	☐	✓	Needed special equipment	
						Needed special knowledge	
						Expected cost	low
Remarks: Widely used for internet services and smartphone apps							

Table 35: Applicability – Monitoring of service use statistics.

Method 10: Monitoring of service use statistics								
Applicability								
Assessment / assurance		Part value chain		Type of service (equipment)		Quality criteria		Type of result
Assurance	✓	Content detection	☐	Equipment	☐	Geographical coverage	✓	quantitative (but service quality indirect, qualitative)
Assessment	✓	Content processing	☐	can indirectly indicate equipment faults		Availability	☐	objective
						Timeliness start	☐	
Event / status		Service provision	✓	Process	✓	Reporting period	☐	in direct
				can indirectly indicate process faults		Timeliness update	☐	results do encompass criteria indirectly only
						Latency	☐	
Event	✓	Service presentation	✓			Location accuracy	✓	
Status	✓					Reporting accuracy	✓	
Offline / online						Classification correctness	✓	
Offline	✓					Event coverage	✓	
Online	✓					Report coverage	✓	
Remarks	Provides quantitative information essential to service provider but only indirect with regard to specific quality criteria							

4.2. Relevance and applicability of the assessment methods

The table below (Table 36:) summarises the purposes for ´which the methods can be applied, the coverage of the methods in the RTTI value chain, their applicability to quality assurance or assessment, applicability to event or status oriented information and assessment of individual pieces or types of equipment or the service process.

Table 36: Summary on applicability of analysed quality assessment methods and practises.

		Objective				Coverage of value chain				Assess ment / assu-rance		Event / status		Type of service / equipment	
		Assessment of service	Acceptance testing	Feasibility / testing new procedure of algorithm	Internal quality control / monitoring	Content detection	Content processing	Service provision	Service presentation	Quality assurance	Quality assessment	Event	Status	Equipment	Process
1	Continuous monitoring of equipment performance and availability	X	X	X	X	X				X		X	X	X	
2	Manual verification of events or conditions	X	X	X	X	X	X	X	X		X	X		X	X
3	Reference testing of data collected	X	X	X	X	X	X	X	X	X	X	X	X	X	X
4	Time-space oriented reference test methods	X			X	X		X			X	X			X
5	Monitoring of data completeness and latency	X			X	X	X	X		X	X	X	X	X	X
6	Regular sampling of mes-sage or data content comp-leteness and correctness				X		X			X		X			X
7	Verification and calibration of traffic / weather conditions prognosis	X	X	X	X		X	X	X	X	X	X	X		X
8	Surveys of perceived quality by users	X			X	X	X	X	X		X	X	X	X	X
9	Collection of direct user feedback	X				X	X	X	X	X		X	X	X	X
10	Monitoring of service use statistics	X						X	X	X	X	X	X		X

Mapping between the quality assessment methods and the elements of the quality criteria for SRTI and RTTI services is provided in Table 37:. The table also indicates the types of results expected with the quality assessment methods.

Table 37: Results provided by analysed assessment methods. X means that the method can be used to address the criteria directly, and x means indirectly.

.		Quality criteria											Type of expected result					
		Geographical coverage	Availability	Timeliness start	Reporting period	Timeliness update	Latency	Location accuracy	Reporting accuracy	Classification correctness	Event coverage	Report coverage	Qualitative	Quantitative	Objective	Subjective	Direct	Results do encompass criteria
1	Continuous monitoring of equipment performance and availability		x			x			x		x			X	X			X
2	Manual verification of events or conditions			X		X		X		X			X		X		X	X
3	Reference testing of data collected			X		X	X	X	X	X	X	X		X	X		X	X
4	Time-space oriented reference test methods			x		x	x	x		x	x			X	X			X
5	Monitoring of data completeness and latency		X	X	X	X				X				X	X		X	X
6	Regular sampling of message or data content completeness and correctness			X		X	X	X	X	X	X		X		X		X	
7	Verification and calibration of traffic / weather conditions prognosis								X	X	X	X		X	X		X	X
8	Surveys of perceived quality by users		x	x		x	x			x	x			X		X		
9	Collection of direct user feedback		X					X	X	X			X			X	X	X
10	Monitoring of service use statistics	x						x	x	x	x	x	(X)	X	X			

5. Conclusions

The Delegated Regulations (EU) No 2015/962 and No 886/2013 contain some implications to commonly describe and document the quality of services and underlying data in the field of Safety-Related Traffic Information (SRTI) and Real-Time Traffic Information (RTTI).

EU EIP (and its antecessor projects) has developed a framework how to commonly describe and document quality in this context.

The presented “Quality Package” contains quality-related definitions and concepts, as proposed and agreed by EU EIP partners for the use in Europe:

- A set of quality criteria in the categories ‘Level of Service” (describing the provision of data) and ‘Level of Quality” (describing the data as such) → see chapter 2.2
- Specific Quality Requirements for individual SRTI/RTTI services → see chapter 3
- A set of proposed Quality Assessment Methods → see chapter 4

The main lesson learned from the EU EIP activity is that any quality definition requires a continuous back-up and validation from stakeholders. For this reason, evidence from conditions and operating environments, as well as expert knowledge of the public and private stakeholders has been involved in the EU EIP quality work.

After several validation and revision steps, this current version of the “Quality Package” presents validated quality definitions, ready to be used in every-day practice by all SRTI and RTTI services and stakeholders in Europe. However, further feedback and expertise from European practitioners may result in future revisions of the “Quality Package”.

Any potential driver for future revisions may be near-future technologies, such as C-ITS. These technologies most likely will have a significant impact on SRTI and RTTI services, e.g. by improving detection and communication of related data. As a result, we can expect significant potentials to improve achievable quality levels. However, these potentials cannot be exactly fully described today. Thus, quality-related effects of near-future technologies have to be carefully analysed and eventually incorporated into this quality framework. In this version, a quality level **** has been defined for Quality Requirements which will be related to near-future technologies, but it has not been specified with concrete values yet.

6. References

Bogenberger,K. 2003a. Qualität von Verkehrsinformationen. Strassenverkehrstechnik, Vol. 47, Issue 10.

Bogenberger, K. 2003b. The Quality of Traffic Information. http://www.trafficforum.ethz.ch/vwt_2003/beitraege/VWT19proceedings_contribution_28.1-28.17.pdf [accessed 20th March 2014]

Bogenberger, K. 2003c. The quality of traffic information.

Bogenberger, K. 2010. A Macroscopic and a Microscopic Model for Measuring the Quality of Traffic Information, presentation slides. http://www.lakeside-conference.at/fileadmin/downloads/LC2010_Presentations/LC2010_Session-5-2_Bogenberger%20Klaus.pdf [accessed 5th January 2016]

Bogenberger, K. and Hauschild, M. 2009. QFCD – A Microscopic Model for Measuring the Individual Quality of Traffic Information. Proceedings of the 16th ITS World Congress, 2009.

Bogenberger, K. and Weikl, S. 2012. Quality Management Methods for Real-Time Traffic Information. Procedia Social and Behavioral Sciences, Vol. 54, Issue 4, pp. 936-945.

Chen, C., Kwon, J., Rice, J., Skabardonis, A. and Varaiya, P. 2003. Detecting errors and imputing missing data for single loop surveillance systems. 82nd TRB Annual Meeting, January 2003, Washington D.C., USA.

Dance, F., Gawley, D., Hein, R., and Kates, R. 2008. Enhancing Navigation Systems with Quality Controlled Traffic Data, SAE Technical Paper 2008-01-0200

Hendriks, T, van der Vliet-Hameeteman, C, 2015a. QSRTI: a method for quality assessment of event based traffic information with a focus on Safety Related Traffic Information. TNO report 2015-1003, 19 February 2015.

Hendriks, T, van der Vliet-Hameeteman, C, 2015b. QSRTI: a method for quality assessment of event based (safety related) traffic information., 22nd ITS World Congress (paper ITS-2050), Bordeaux, France, 5–9 October 2015.

Känsäkangas, T. Quality Assessment Methods for Traffic Incident Information, Masters Thesis, Aalto University, Finland, 14 September, 2015.

Li, H. and Bertini, R. K. 2010. Comparison of Algorithms for Systematic Tracking of Patterns of Traffic Congestion on Freeways in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board, No. 2178, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 101–110

Lohoff, J., Ansorge, J., Rystrøm, L., Kulmala, R., Öörni, R., Hendriks, T. Kusters, M and Dumitrescu, S. 2015. Framework Guidelines for Data and Service Quality Requirements, version 1.0 23rd March 2015. Final Report of EIP Sub-Activity A3.2.

Kulmala, R., Öörni, R., Dean, S., Hendriks, T., Tomas, V., Nemtanu, F., Laine, T., Rystrøm, L., Hovestad, M., Ansorge, J. and Schirokoff, A. 2015. Proposal for quality assessment of ITS services on the TERN, version 1.01, 11th March 2015.

Kulmala, R., Öörni, R., Laine, T., Rystrøm, L., Roberg, C., Dumitrescu, S., Ansorge, J. and Narroway, S. 2015. Quality criteria, requirements and assessment test plan, version 1.0, 23rd May 2015. Deliverable of EIP+ 3.1 Testing and Validating of the Quality Recommendations and Results for Traffic Information from EIP.

Limber, M., Drobot, S. and Fowler, T. 2010. Clarus quality checking algorithm documentation report. U.S. Deparment of Transportation, Washington D.C., USA. http://ntl.bts.gov/lib/38000/38500/38545/TOPR2_508_FHWA-JPO-11-075.pdf [accessed 5th January 2016]

Lu, X-Y. 2008. Faulty loop data analysis/correction and loop fault detection. Proceedings of the 15th World Congress on Intelligent Transport Systems, November 16-20 2008, New York, New York, USA.

Lu, X-Y., Kim, Z., Cao, M., Varaiya, P., and Horowitz, R. 2010. Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data. California PATH Research Report UCB-ITS-PRR-2010-5, University of California, California, USA. http://its.berkeley.edu/sites/default/files/publications/UCB/2010/PRR/UCB-ITS-PRR-2010-5.pdf [accessed 5th January 2016]

Mediamobile 2014. Mediamobile – our values.

http://www.mediamobile.com/index.php/en/our-values [accessed 5th January 2016]

Penttinen, M., Innamaa, S., Pilli-Sihvola, E., Aittoniemi, E. and Rämä, A. 2014. Use and Impacts of Travel Time Information in Finland. Proceedings of the 10th European Congress on ITS, 16-19 June 2014, Helsinki, Finland.

Rystrøm, L. Quality of Safety Related and Real-time Traffic Information, Practical Guidelines. Version 0.9. EU EIP SA 4.1. 8 December 2017.

U.S. DOT. 2014. Success stories, Clarus. http://www.its.dot.gov/clarus/ [accessed 5th January 2016].

Wilks, D. 2005. Statistical methods in atmospheric sciences. Academic Press, Burlington, MA, USA.

Wiltschko, T. 2004. Sichere Information durch infrastrukturgestütze Fahrassistenzssysteme zur Steigerung der Verkehrssicherheit an Straßenknotenpunkten. Dissertation at Stuttgart University