Quality of Safety Related and Real-time Traffic Information

Evaluation of quality tests

Final report

EU EIP SA 4.1: Determining Quality of European ITS Services

Version: FINAL

Date: 12th June 2018

Co-financed by the European Union
Connecting Europe Facility

1. Document Information

Authors

Name	Organisation
Tomi Laine	Strafica Ltd (FI)
Risto Öörni	VTT (FI)
Peter Lubrich	Federal Highway Research Institute BAST (DE)
Leif Ryström	Vejdirektoratet (DK)
Michel Kusters	Rijkswaterstaat (NL)
Gema Leiro Garcia	ICEACSA (ES)
Jacqueline Barr	IBI Group (UK)
Radu Milea	CESTRIN (RO)
Stephanie Cheung	Hessen Mobil - Straßen- und Verkehrsmanagement (DE)
Risto Kulmala	Liikennevirasto (FI)
Anna Schirokoff	Trafi (FI)

Distribution

Date	Version	Dissemination

2. Preface

The European ITS Platform (EIP) project running from 2013 to 2015 developed the quality framework for real-time and safety-related traffic information as described in the Delegated Acts b and c of the ITS Directive. The project's output was the initial target values for collecting experiences and validation in future projects. The first validation was performed in EIP+ project in 2015, which resulted in the publication of the Quality package for safety related and real-time traffic information services.

In the EU EIP sub-activity 4.1 Determining Quality of European ITS Services the validation work regarding the quality framework was continued, and this is the deliverable of the quality validation task within the activity. Several changes to the initial target values were proposed to increase the feasibility of the requirements across Europe and to provoke ambition for further development of service quality for end users. Due to the successful validation of most of the quality criteria the proposed new requirements are called "validated target values".

There were seven participants to the quality tests, one from each of the following countries: Germany, The Netherlands, The UK, Denmark, Romania, Spain and Finland. The test set-up, involving various partners in the ITS field and various approaches, allowed to cover many of the defined quality criteria and assessment methods, so a solid ground for further improvements of the quality proposals is given.

TABLE OF CONTENTS

1. Document Information **2**

2. Preface **3**

3. Introduction **6**

3.1. Context 6

3.2. Goal of the quality tests 7

4. Performed tests **13**

4.1. Road works information (Germany) 13

4.2. Real-time traffic event information (Denmark) 20

4.3. Weather information (Romania) 21

4.4. Travel time information and road weather prognosis (UK) 24

4.5. Travel time information (Finland) 27

4.6. Safety-related traffic information (The Netherlands) 30

4.7. Safety-related information for reduced visibility (Spain) 36

5. Evaluation of the results **38**

5.1. Quality criteria for SRTI and RTTI 38

5.2. Quality requirements for SRTI and RTTI 41

5.3. Quality assessment methods 47

6. Proposal for the validated quality recommendation for SRTI and RTTI 53

6.1. Quality criteria for SRTI and RTTI 53

6.2. Quality requirements for SRTI and RTTI 53

6.3. Quality assessment methods 61

6.4. Practical examples and interpretation of the requirements 62

7. Conclusions 65

8. References 67

3. Introduction

3.1. Context

The purpose of the EU EIP sub-activity 4.1 “Determining Quality of European ITS Services” is to propose quality criteria, minimum requirements and assessment methods for all priority actions of the EU ITS Directive. The current EU EIP project focuses on multimodal travel information and planning services (priority action a), intelligent truck parking (e), and cooperative ITS. However, in the early phases of the EU EIP project, it was decided to also further validate the earlier quality proposals for real-time traffic information services (RTTI, priority action b) and safety related traffic information services (SRTI, priority action c), and simultaneously work with the quality proposals for other mentioned services. Therefore another validation task regarding SRTI and RTTI quality requirements was added to the workplan of EU EIP sub-activity 4.1, defined as “Task 3.0 – Quality Test for RTTI and SRTI”. The idea of the validations is that the quality criteria, minimum requirements and assessment method proposals are validated by partners in real-life quality assessments of the service deployments, and when necessary, the proposals are adapted to the findings of the validation to improve their viability for European use.

The quality tests summarised in this report are a continuation for the validation work started in the EIP+ sub-activity 3.1 “Testing and Validating of the Quality Recommendations and Results for Traffic information from EIP”. On the first round of validations in EIP+ project the main learning was that obtaining good quality assessment data from the existing ITS systems is many times difficult and that further validation studies of the quality criteria for RTTI and SRTI need to be carried out both as national studies and as actions in the EU EIP. No justification was found to alter the initial quality requirements established already in the EIP project.

The Quality Deliverables from EIP+ sub-activity 3.1. formed the starting grounds for the new round of validations, namely the “Quality package for safety related and real-time traffic information services” and the report “Validated quality criteria, requirements and assessment methods for safety related and real-time traffic information services”.

In EU EIP sub-activity 4.1 it was decided to continue using the mentioned “Quality package”, and to update the package with the latest changes to the quality proposals, namely descriptions of the quality criteria, requirements and assessment methods, as agreed in the sub-activity 4.1 working group.

3.2. Goal of the quality tests

The goal of the performed quality tests was to test the quality proposals (quality criteria, minimum requirements and assessment method descriptions) under real-life operating conditions in various countries involved in the EU EIP project. The purpose was to ensure that the quality criteria are applicable and the quality requirement levels, especially the minimum level requirement are initially set to a level that can be reached in most of the member states using reasonable amount of efforts. In addition, the purpose was to validate the assessment methods and their descriptions and to ensure that they support the measurement of the presented criteria with sufficient amount of detail. The test participants were guided to pay increased attention in encountered challenges regarding the applicability of the quality criteria and assessment methods.

The quality proposals validated in this study, based on further elaborations of EIP+ deliverables, are presented in the following tables.

*Table 1: **Initial Target Values** for the Level of Quality (parts of RTTI: Real-time event information / Traffic conditions information (TCI) / Weather information) for continuative evaluation after EIP+*

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start (95 %)	**Best effort**[1]	Acceptance[2] after first detection < 10 min[1] Time between occurrence and first detection: Best effort[1]	Detection & acceptance[2] < 5 min after event occurrence	Detection & acceptance[2] < 3 min after event occurrence
Timeliness update/end (95%)	**Best effort**[1]	Best effort[1]	Detection & acceptance[2] < 10 min after event change/end	Detection & acceptance[2] < 5 min after event change/end
Latency (content side) (95 %)[3]	**< 10 min**[1]	< 5 min[1]	< 3 min	*< 2 min*
Location accuracy (95%) - Area	**Administrative region (TCI: n/a)**	Geographic area; 10 km accuracy *(TCI: n/a)*	Geographic area; 5 km accuracy *(TCI: n/a)*	*Geographic area; 2 km accuracy (TCI: n/a)*
Location accuracy (95%) - Road	**link between Intersections**[4]	< 10 km or link between Intersections[4]	< 5 km[4]	*< 2 km*[4]
Classification correctness	**> 85%**	> 90%	> 90%	*> 98% (TCI: > 95%)*
Event coverage	**Best effort**	Best effort[5]	> 80% of all occurring events	*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an organisation's quality policy*

[3] *Counts for 95% of the analysed data address the latency measurements at the start and at the end of events.*

[4] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.*

*Table 2: **Initial Target Values** for the Level of Quality (part of RTTI: Travel Time information) for continuative evaluation after EIP+*

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Reporting period	**Best effort**	5 min	1 min	
Timeliness update (95%)	**Best effort**	< 5 min	< 2 min	
Latency (content side) (95%)	**< 10 min**	< 5 min	< 2 min	*< 1 min*
Reporting accuracy (95%)	20%	10%	5%	
Classification correctness	**90%/80%**	95%/80%	98%/80%	
Event coverage	**Best effort**	>90%	>97%	

*Table 1: **Initial Target Values** for the Level of Quality (part of SRTI: Wrong Way Driver) for continuative evaluation after EIP+*

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start (95%)	**Best effort**[1]	Acceptance[2] after first detection < 5 min[1] Time between occurrence and first detection: Best effort[1]	Detection & acceptance[2] < 2 min after event occurrence	Detection & acceptance[2] < 1 min after event occurrence
Timeliness update/end (95%)	**Best effort**[1]	Best effort[1]	Detection & acceptance[2] < 5 min after event change/end	Detection & acceptance[2] < 2 min after event change/end
Latency (content side) (95%)[3]	**< 3 min**[1]	< 2 min[1]	< 1 min	*< 1 min*
Location accuracy (95%) - Area	**Administrative region (TCI: n/a)**	Geographic area; 10 km accuracy *(TCI: n/a)*	Geographic area; 5 km accuracy *(TCI: n/a)*	*Geographic area; 2 km accuracy (TCI: n/a)*
Location accuracy (95%) - Road	**affected road segment not longer than 50 km or road number with destinations (city names) nearby**[4]	link between Intersections[4]	link between Intersections[4]	*< 5 km*[4]
Classification correctness	**> 75%**	> 85%	> 90 %	>90 %
Event coverage	**Best effort**	Best effort[5]	> 80% of all occurring events	*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an organisation's quality policy*

[3] *Counts for 95% of the analysed data address the latency measurements at the start and at the end of events.*

[4] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.*

[5] *Definition of criterion and value need further examination.*

*Table 2: **Initial Target Values** for the Level of Quality (parts of SRTI: all SRTI events/conditions, except wrong way driver) for continuative evaluation after EIP+*

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start (95%)	**Best effort**[1]	Acceptance[2] after first detection < 10 min[1] Time between occurrence and first detection: Best effort[1]	Detection & acceptance[2] < 5 min after event occurrence	Detection & acceptance[2] < 3 min after event occurrence
Timeliness update/end (95%)	**Best effort**[1]	Best effort[1]	Detection & acceptance[2] < 10 min after event change/end	Detection & acceptance[2] < 5 min after event change/end
Latency (content side) (95 %)[3]	**< 10 min**[1]	< 5 min[1]	< 3 min	*< 2 min*
Location accuracy (95%) - Area	**Administrative region**	Administrative region	Geographic area; 10 km accuracy	*Geographic area; 2 km accuracy*
Location accuracy (95%) - Road	**link between Intersections**[4]	< 10 km or link between Intersections[4]	< 5 km[4]	*< 2 km*[4]
Classification correctness	**> 85%**	> 90%	> 90%	*< 2%* *(TCI: < 5%)*
Event coverage	**Best effort**	Best effort[5]	> 80% of all occurring events	*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an organisation's quality policy*

[3] *Counts for 95% of the analysed data address the latency measurements at the start and at the end of events.*

[4] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.*

There were seven participants to quality tests, one from each of the following countries: Germany, The Netherlands, The UK, Denmark, Romania, Spain and Finland. The test set-up, involving various partners in the ITS field and various approaches, allowed to cover many of the defined quality criteria and assessment methods, so a solid ground for further improvements of the quality proposals is given.

After the validation tests have been performed and evaluated, the actual validation on the European scale will be carried out in 2018 in targeted validation workshops (co-arranged with TISA, CEDR and ASECAP). The consolidation of validation findings will also enable recommendations for improving the quality proposals.

This report contains:

- a description of the individual tests, including the test settings, results and implications on the RTTI and SRTI quality definitions so far (section 4),
- a consolidated evaluation of the results, looking specifically on possible effects on quality criteria, requirements and assessment methods (section 5),
- a proposal for the validated quality recommendation for SRTI and RTTI, reflecting individual quality criteria, requirements and assessment methods (section 6) and
- summarizing conclusions (section 7).

4. Performed tests

4.1. Road works information (Germany)

4.1.1. TEST SETTING

The quality test implemented by the German partners aimed to check the correctness of published information on road works. For this reason, public information from Traffic Information Centres was checked against comparable information coming from a private service provider.

In particular, the following data sources were analysed:

1. Three data feeds (DATEX II format) of road work information from road operators provided via the Mobility Data Marketplace (MDM - German National Access Point (NAP))
    a) RTTI events: Long-term road works in Hessen (part of the national information platform "Baustelleninformationssystem")
    b) SRTI: Short-term road works in Hessen (proprietary publication of the state Hessen)
    c) RTTI events: Road works in Nordrhein-Westfalen (proprietary publication of the state Nordrhein-Westfalen)
2. Live-Traffic Messages from service provider TomTom (DATEX II format) as a ground truth.

The tested quality criteria were **Location Accuracy**, **Classification correctness** and **Event Coverage**. The method chosen was Reference testing of data collected.

Road work information from public road operators in Germany are known to have some inconsistent data quality nowadays for various reasons:

- very different procedures in processing road work information in the different organisations involved
- different understanding/interpretation for DATEX II data fields
- discrepancy between "planned" and "real-life" road work situations (in some cases, only the "planned" road work situation is published by the road operators; "real-life" situations are not always tracked and published in real-time).

This unsatisfactory situation is addressed by private service providers by validating and "optimising" the public road work information. In case of TomTom, data from road operators is validated based on floating car data speeds coming from the company's data sources. The company states that data quality is improved this way.

In this context, it was assumed that TomTom data could be handled as "ground truth".

The quality test consisted of the following subsequent analysis steps:

1. Record information from public road operators (sources no. 1a, 1b, 1c) from the MDM for a fixed point of time (i.e. no analyses over longer time-periods were done).
2. Simultaneously, record traffic feed from TomTom (source no. 2), for the same fixed point of time.
3. In all recordings: Identify individual road work events, classify identified road work events (in terms of time, location, type of road work etc.).
4. Match identical road work events in comparable data sources (MDM and TomTom recordings); i.e. identify information that relates to the same road work event, using geospatial matching algorithms and pre-defined confidence levels.
    - For matched events (= events identified in both data sources): Compare location information for starting and ending point of reported road works, determine **Location accuracy** for this particular event.
    - For non-matched events (= events only identified in one data source): Try to identify the reason for the non-existence in the other source; determine **Classification correctness** and **Event Coverage** for this particular event.
5. Consolidate Quality Criteria for the following data sets:
    - TomTom vs. MDM publication "Long-term road works in Hessen"
    - TomTom vs. MDM publication "Short-term road works in Hessen"
    - TomTom vs. MDM publication "Road works in Nordrhein-Westfalen".

An automated JAVA programme was written to conduct the steps 1 till 4. Basic Office software was used for the analysis in step 5.

Below are some details on the calculation of the quality criteria:

Location accuracy was calculated only for road work messages that were matched between the MDM publications and the TomTom source. It was based on comparing individual location identifiers (using LCL definitions), i.e. by checking if the location identifier (e.g. "junction XY") in one data source is the same one as in the other source. In addition, the relative distance of coordinates of a location in both sources (expressed in meters) was checked. Location Accuracy was determined both for the start point and the end point of each message. The reason for this is that, unlike other RTTI events (such as incidents), road works usually stretch over a section of a roadway. Consequently, road works are located by a start point and by an end point.

Classification correctness was calculated with the following formula:
$1 - [(f_p / f_p + r_p) * 100\%]$
with: f_p = number of "false positive" messages ("ghosts")
r_p = number of "correct positive" messages ("hits")
"Ghosts" are road work messages that appear in the MDM publications but not in the TomTom source.
"Hits" are road work messages that are matched between the MDM publications and the TomTom source.

Event Coverage was calculated with the following formula:
$(r_p / r_p + f_n) * 100\%$
with: r_p = number of "correct positive" messages ("hits")
f_n = number of "false negative" messages ("missies")
"Hits" are road work messages that are matched between the MDM publications and the TomTom source.
"Missies" are road work messages that appear in the TomTom source but not in the MDM publications.

4.1.2. RESULTS

The results from the quality tests are presented in the Table 3 to Table 5.

Table 3: Results from the German quality tests, data sets: TomTom vs. MDM publication "Long-term road works in Hessen (RTTI)"

Criteria	Filtering condition	Value	Quality level
Location accuracy(95%) - Road	See remarks on the calculation of the quality criteria above.	*For start point:* 4324 m *For end point:* 7403 m	*For start point:* *** Advanced *For end point:* ** Enhanced
Classification correctness		38 %	none
Event Coverage		36 %	Best Effort

Table 4: Results from the German quality tests, data sets: TomTom vs. MDM publication "Short-term road works in Hessen (SRTI)"

Criteria	Filtering condition	Value	Quality level
Location accuracy(95%) - Road	See remarks on the calculation of the quality criteria above.	The system currently (02.02.18) does not deliver any information, assuming a technical failure. Thus, no analyses can be run.	
Classification correctness			
Event Coverage			

Table 5: Results from the German quality tests, data sets: TomTom vs. MDM publication "Road works in Nordrhein-Westfalen (RTTI)"

Criteria	Filtering condition	Value	Quality level
Location accuracy(95%) - Road	See remarks on the calculation of the quality criteria above.	*For start point:* 4909 m *For end point:* 5008 m	*For start point:* *** Advanced *For end point:* ** Enhanced
Classification correctness		28 %	none
Event Coverage		77 %	Best effort

To get a better picture on the quality criterion **Location accuracy**, a more detailed analysis for the data set " Road works in Nordrhein-Westfalen" was conducted, looking more specifically on the fulfilment of each of the defined quality levels *Basic **, *Enhanced ***, *Advanced ****, *and *****. For the sample size of 120 messages, the following diagram shows of how many these messages fulfil the defined quality levels.

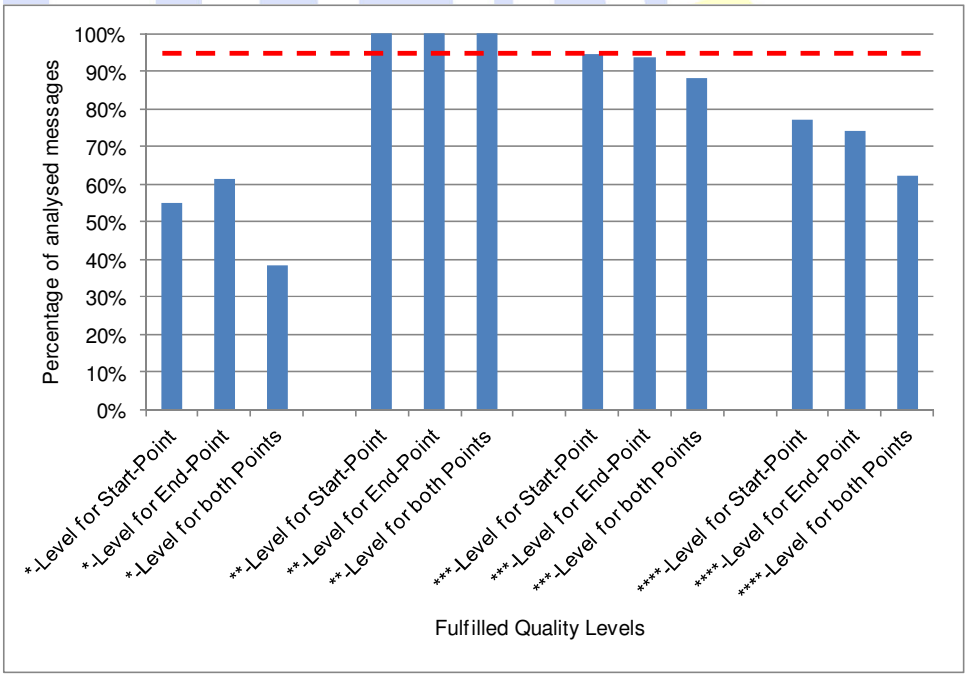


*Figure 1: Detailed analysis of the German quality tests regarding **Location accuracy***

4.1.3. IMPLICATIONS

Implications on **Location accuracy**:

- The *Enhanced *** level (required accuracy <10 km) can be easily fulfilled by all analysed data sets.

- The *Advanced **** level (required accuracy <5 km) is narrowly missed. It seems that the *Advanced **** level could be also fulfilled easily.

→ **It is recommended to define the required accuracies for *Enhanced *** and *Advanced ***** levels stricter than 5 and 10 km.** It is assumed that road work information must be somewhat more accurate for the following reasons:

  - When a reported road work is located more than 5 km apart from the real position, there must be something wrong with the information source, or even a very different road work is meant by the message.
  - For a road user, a road work message with an error above 5 km is useless, as he wishes to adopt his driving behaviour and also his route choice in close distance to the real road work location.

- For the *Basic ** Level (location required within the correct "link between intersections"), it is found that it cannot be reached for the analysed data sets. This can be explained with the inconsistent georeferencing of the data sets (see "technical challenges" below). Because of the different geospatial granularity of the compared data sets, the road work locations very often showed offsets when looking at the road links, see the following example. In this example, both the start and the end point of the road work have a small offset, leading to a situation where their reported location is not exactly between the same intersections. Even if applying a tolerance of 100 metres, as suggested in the latest *Quality Package*, the *Basic ** level could not be reached.

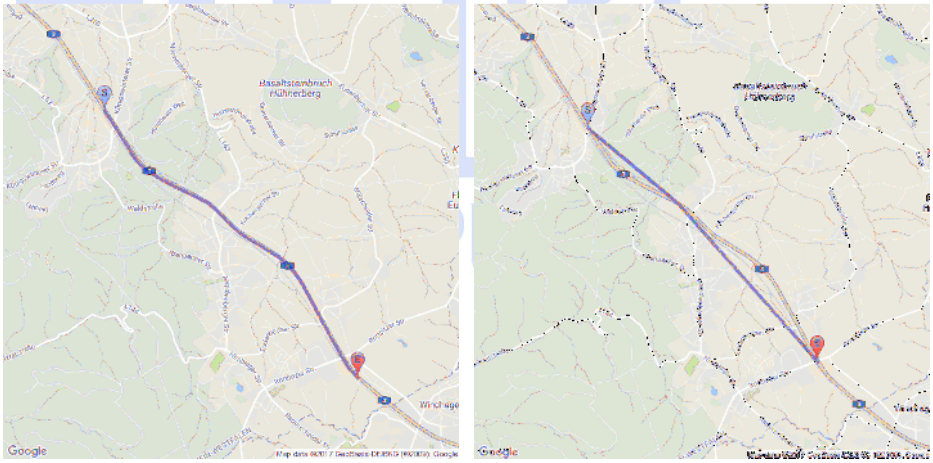


Figure 2: Example for different location for one road work message based on two compared data sets

→ **It is recommended to rethink the "link between intersections" concept as a *Basic ** level for this type of assessment method**, as it may lead to wrong judgment about location accuracy in case of inconsistent geoferencing systems. Instead, a distance-based level (comparable to the *Enhanced *** and *Advanced **** levels) could be introduced.

Implications on **Classification correctness**:

- The determination of this criterion depends on the number how many messages could be matched (or not matched) between the MDM and TomTom data sources. For the matching, there are challenges with the interpretation of “one road work event” and “one road work message” in both data sources. With the analysis method applied here, Classification correctness was determined to 30-40%, far below the *Basic ** level of minimum 85%. However, this is related to the mentioned interpretation problems. Thus, further findings for the validation of this criterion cannot be given yet.

Implications on **Event Coverage**:

- Again, the determination of this criterion depends on the number how many messages could be matched (or not matched) between the MDM and TomTom data sources. In this case, no real comparison could be done as the MDM sources cover some of the road network (mostly motorways and Federal roads) and TomTom covers the entire road network. A special filtering was applied to select only the correspondent MDM messages. However, a verification of this filtering could not be done. The resulting numbers of 36 % and 77 % are not trustworthy to give findings for the validation of this criterion.

Looking at the assessment method applied for this quality test, a huge challenge was seen in the difference of the “data worlds“ that had to be brought together:

- Road operator data (here: three different MDM publications):
  - originating from various sources
  - influenced by different organisational procedures
  - mainly produced for a planning perspective
- Private service provider data (here: TomTom traffic feed) :
  - validated data using own FCD resources
  - mainly produced for a road user, real-time perspective.

It became quickly obvious, that messages describing the current road work situation are quite inconsistent in both “data worlds”, creating some technical challenges within the quality analyses:

- Matching algorithms to identify identical road work events:
  The matching was based on geospatial analyses of reference points contained in the data sets, such as mapping with the closest network nodes as well as the calculation of geographic distances of reference points. Depending on the set-up of the matching algorithms, two events may be declared as “matched” or as “non-matched”, eventually resulting in different quality parameters.
- Georeferencing:
  As different georeference systems were used in the different data sources, some of the data had to be converted first to be comparable in a geographic sense. However, the granularity of the geodata sometimes remained incomparable. TomTom is referencing road works to specific start and end node coordinates. On the other hand, road work information from German road operators is usually

located via the Location Code List[1] (LCL). If there is no specific "offset"-information attached, a road work would be mapped to specific locations determined by LCL, see the example below (Figure 3). There may be some issues for assessing **Location Accuracy** based on the "link between intersections" concept, when using inconsistent georeferencing (see above).

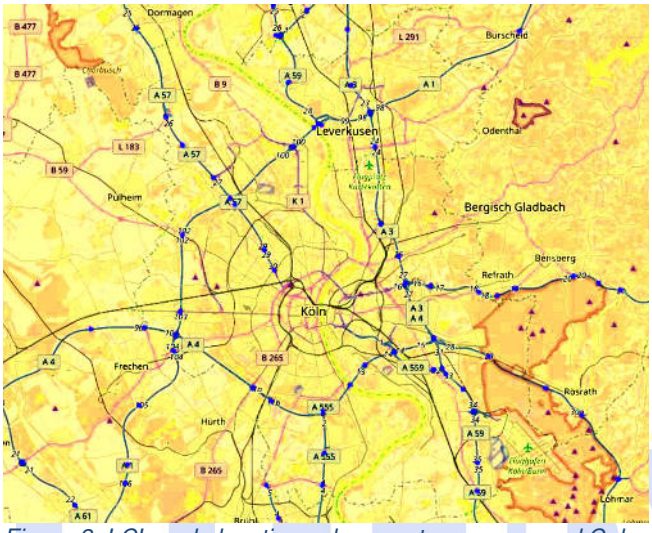


Figure 3: LCL node locations along motorways around Cologne, Germany

- The interpretation of "one road work event" and "one road work message": Due to different understanding how to describe road work events, different data sets show some systematic discrepancies. For example, a road operator may publish an entire motorway section in which a road work "happens" in a wider sense (including speed-restriction zones upstream of the road works), while a service provider would only publish the specific section where a capacity/speed drop is actually influencing the driver's travel time. In other cases, longer-period road works (e.g. for a more than one year) are declared as "normal state" by service providers and included in their "regular" navigation maps without a specific message, while a road operator would publish a permanent message here. In other cases, some road works were reported redundantly (via multiple messages) in the MDM data sources, leading to problems when "matching" with TomTom data. All these cases make a comparison of messages difficult.

Another issue is the ground truth assumption made for this quality test. There are some hints that road work information from the service provider reflect the current traffic situation in a more complete and consistent way, comparing to information from road operators. However, no scientific evidence has been given that road work information from the service provider has a higher quality or even can be declared as "ground truth".

[1] http://www.bast.de/DE/Verkehrstechnik/Fachthemen/v2-LCL/lcl_node.html

The challenges described above have to be considered when different “data worlds” are applied for quality analyses. It is understood that data sources from service providers (FCD etc.) are quite new to road authorities (either for data quality analyses or other tasks of traffic management) and that there are only few experiences or standards. Therefore it is difficult to find an agreement today how to deal with that kind of data on a pan-European level. Further analyses and potentially the development of guidelines how to use service providers’ data for road authorities are highly recommended.

4.2. Real-time traffic event information (Denmark)

4.2.1. TEST SETTING

In Denmark the Vejdirektoratet performed a quality test for the DRD traffic information service. The purpose was to measure the following quality parameters: Latency (content side, start), Latency (content side, end), Location accuracy and Classification correctness.

The study sample consisted of 50 accidents that occurred during week 35, the last week of August 2017 and were acknowledged by the Traffic Management Centre (TCC). The data was extracted from the IT-system and was manually handled in Excel spreadsheet. No specific ground truth was used.

4.2.2. RESULTS

The results from the quality tests are presented in the following table.

Table 6: Results from the Danish quality tests for RTTI event information

Criteria	Filtering condition	Value	Quality level
Latency (content side, start)	none	3/50 events with latency over 5 minutes	* Basic
Latency (content side, end)	none	all event below 5 minutes	** (level *** not tested?)
Location accuracy	none	all correct	***
Classification correctness	none	all correct	***

4.2.3. IMPLICATIONS

The Danish participant indicated that the dataset gives an indication of the quality parameters, but 50 events is a too small sample to give accurate results. The sample consists of only accidents (RTTI - Event information and SRTI - Unprotected accident area), but because in the traffic management centre all SRTI categories and RTTI events are handled in (almost) the same way, it can be assumed that the results apply to all event types and categories.

4.3. Weather information (Romania)

4.3.1. TEST SETTING

The Romanian partners implemented quality tests regarding RTTI Weather information. For these tests, two weather stations, installed on the A1 Sibiu - Orastie highway to a 16 km section, were used. The weather stations are monitored in Saliste Traffic Information Centre (TIC). This highway section is located in a hilly area where there are quite a few meteorological variations over short periods of time. Initially, these tests had to be carried out over a 72-hour period (between 26.10.2017 at 07:00 till 29.10.2017 at 07:00), but due to a failure of the power supply network of the weather equipment, the period of the test was increased to 84 hours (until October 29, 2017 at 19:00). For monitoring and managing weather equipment and all ITS equipment installed on this highway section, a Sielte Roads application is used.

The goal of the test was to find out, how the equipment are being checked (Report Coverage), what is the time span for refreshing / updating the equipment status (Reporting Periods), what is the time period needed from receiving a weather alert until it is accepted by the operator (Timeliness update) and what is the time period from accepting an event by the operator until the message is displayed on the VMS (Latency).

The monitoring and management application that integrates all ITS equipment (including weather stations) has the ability to continuously monitor equipment status, generate alarms automatically when an event occurs outside preset parameters, and also generate alarms in the case of communication failures or any technical dysfunctionalities of the equipment. In order to verify the authenticity of the signals automatically transmitted and their validation, operators use any other ITS equipment mounted in that area that could provide additional information (traffic cameras, PTZ cameras, weather stations, etc.) as well as any other information from authorities and state institutions (National Meteorological Administration, Police, etc.).

4.3.2. RESULTS

The results from the quality tests are presented in the following table.

Table 7: Results from the Romanian quality test for weather information.

Criteria	Filtering condition	Value	Quality level
Report coverage (in case of technical failures)	sample size only 2 events: the technical failure and the restoration of functions	100 %	sample size too small for evaluation
Timeliness (update) 95%	seven events (4 for rain, 3 for wind), measured time taken to accept the alarm	5-30 seconds	** Enhanced, not possible to measure *** Advanced
Latency 95 %	two individual events, one for rain and one for wind	rain event 1 min 30 s wind event 40 sec	**** Sample size too small for statistically sound results

Turning off the power supply of the ITS equipment installed on this highway sector did not affect the proper operation of the Sielte Roads application for monitoring and managing the equipment (the application worked properly generating alarms of loss / restoration of communication with the monitored equipment). After reporting the loss of communication with weather stations, verification of this was done by the TIC operator in 10 seconds. Verification of connection restoration with weather stations was performed by the TIC operator in 12 seconds.

For the timeliness test, there were registered 7 alerts (4 for rain and 3 for wind). The average acceptance time of these alerts was 12 seconds (ranging from 5 seconds to 30 seconds). The results pass the level ** Enhanced requirement (Best Effort). However, it is not possible to test level *** Advanced requirement, because there is no data about the time taken to detect the event.

It is mentioned that in addition to weather stations, low visibility caused by fog or heavy rainfall are also detected by CCTV cameras mounted from 2 km to 2 km, used for highway incident detection (these generate alarms in case low visibility).

Regarding the test for latency, it should be noted that currently NAP is not implemented in Romania. In order for Latency to be determined, the moment of publication of the information was considered the moment when the information message was displayed on the VMS. In the case of rain alerts, the information message was displayed in 1 minute and 30 seconds after the alert was accepted by the TIC operator. In the case of wind alert, the information message was displayed in 40 seconds after the alert was accepted

by the TIC operator. All other alerts were considered status checks for initial alerts. The average time needed to display the 2 messages was 1 minute and 5 seconds.

4.3.3. IMPLICATIONS

No direct implications regarding the quality criteria or requirements.

4.4. Travel time information and road weather prognosis (UK)

4.4.1. TEST SETTING

The UK partners implemented two quality tests, one for travel time information and one for RTTI weather information.

1. TRAVEL TIME STUDY

The aim of the travel time quality study was to analyse the quality of actual and predicted journey times for the Traffic Scotland Journey Time System (JTS) against ground truth collected with probe vehicles. The JTS collects journey time data from various sources (ANPR, Bluetooth, Trafficmaster data, loop-based algorithms), processes it (raw data checks, transformation to common link structure, merges to 5 min Fused Data, compares to a normal journey time for that day/time/link) and predicts the journey time for that link for the next time period.

Probe vehicles equipped with GPS logger were used to collect baseline data. Probe vehicles equipped with GPS logger provided data at 5 second intervals. Vehicles drove set routes covering peak am, peak pm, and inter-peak times. The number of runs carried out varied depending on the traffic conditions and length of route. The driver of the probe vehicle had to mimic or match the speed of the traffic stream for a given roadway. One or two vehicles were deployed in each exercise and with surveys lasting 1–3 days depending on the length and complicity of the travel time routes configured in the journey time system. JTS data for matching time periods was extracted from the JTS database, aligned and compared in excel spreadsheet.

2. ROAD WEATHER PROGNOSIS

UK Met Office National Severe Weather Warning Service (NSWWS) warns of the impacts caused by severe weather. Significant historic winter weather events were reviewed from between 2012 and 2017. The Traffic Scotland database was queried to obtain details of any NSWWS weather warnings received from the Met office during these periods. A number of events were manually selected based on the type of weather events forecasted, the number of weather warnings issued, and the duration of the event. For the purpose of this study we were interested in warnings relating to ice, snow, winds, and rain; events that lasted for more than 4 hours and were applied to 3 or more geographic regions.

From the weather events across the time period 41 weather warnings were progressed for analysis.

For each weather warning the Visalia weather station data was extracted for each weather region that was affected by the warning, for the time period the warning was active and for 6 hours before and after the forecasted time. For weather conditions relating to low temperatures and high winds maximum and minimum values were set. The weather station data output included the number of weather stations in the region recording the weather parameters if interest, and maximum / minimum values recorded.

As weather station data is collected in 15 minute intervals, typically 1100 to 4900 data entries for each event per region in the contingency tables were included.

4.4.2. RESULTS

1. TRAVEL TIME INFORMATION

The results from the quality tests are presented in the following table 6.

Table 8: Results from the UK quality tests for Travel Time Information

Criteria	Filtering condition	Value	Quality level
Reporting accuracy (95%) for Actual travel times	link level study	average from all links 17% (range 12-21%)	* Basic
	Route level study	10%	** Enhanced
Reporting accuracy (95%) for Predicted travel times	link level study	average from all links 22% (range 13-27%)	below basic level
	Route level study	10%	** Enhanced

On link level the average reporting accuracy (not an EIP quality criteria) was 6.8% for actual travel times and 6.2% for the predicted travel times. For Transport Scotland actual and predicted journey times, in average, within 10% of the ground truth are considered acceptable. Hence, the results proved satisfactory.

Regarding the reliability of the study results, there are some known difficulties in calculating accurate link journey times from GPS data for links that are very short or that run very close to other roads. Due to financial and time constraints the sample sizes are low, but this method has been suitable for Transport Scotland to date. The partners responsible for the study acknowledge that on some occasions the differences between actual calculated / predicted and probe drive times can be attributed to the dynamic nature of the traffic conditions for the particular lane, time etc., when using limited number of runs. This source for error can be found particularly on short links and routes. Therefore it can be recommended to use the route level results as the primary results, since the measurement differences tend to smooth out when calculated over several links.

2. ROAD WEATHER PROGNOSIS

The results from the quality tests are presented in the following Table 9.

Table 9: Results from the UK quality tests for Road weather prognosis

Criteria	Filtering condition	Value	Quality level

classification correctness	Rain events	93%	****
	Ice events	84%	below basic level
	Snow events	51%	below basic level
Location accuracy		at least one weather station in the geographic area reported the forecast conditions in the predicted time period	how to interpret?
Timeliness start	Data logged only per hour	at least 1 weather station reporting the forecast conditions in the start hour in the cases reviewed.	
Timeliness end		particularly for snow events there was disparity between the end time of the forecast and measured conditions, with the measured conditions ceasing before the predicted event end.	

Across all metrics calculated the results for ice, rain and wind warnings showed better results than for snow events. This may reflect the difficulties in accurately predicting snow events and also collecting accurate snow measurements from the road side equipment.

In general there were high false alarm ratio and bias scores. This probably reflects the preference to "over forecast", erring on the side of caution and giving the worst forecast to a wide area to ensure that drivers receive adequate warning and are prepared for bad weather. This is also a consequence of applying forecast to quite large and disparate geographic areas.

4.4.3. IMPLICATIONS

In the current Quality Package there are specific quality metrics presented for traffic and weather prognosis. These are presented in Table 28 of the Quality Package. They seem to be highly practicable and provide more detailed understanding of the quality of the prognosis, e.g. tendency to over-/underpredict events. However, there is no description of how these metrics tie into the EIP quality criteria and requirements, i.e. is "proportion correct" equivalent to "classification correctness".

4.5. Travel time information (Finland)

4.5.1. TEST SETTING

The purpose of the Finnish study was to study the quality of **RTTI: Travel Time information** for a "connected" induction loop system that was installed for a field pilot. The tested data was acquired from the system called MAVE-S that was installed by the provider AVE Ltd on one motorway stretch on highway 1 in Espoo, Finland. AVE Ltd implemented the system, performed the collection of data and calculated the travel times which were provided to Strafica Ltd for validation. VTT Technical Research Centre of Finland Ltd installed a licence plate registration system (LPRS) system to the same stretch to collect ground truth data for reference testing. Both systems had earlier been used for the same purpose numerous times and the installation staff was highly experienced. Both parties delivered their travel time measurements (1 minute medians for vehicles leaving the measurement link) from four working days that included two severe congestion periods. Clocks were synchronised.

Measurement period of 4 working days resulted in a sample of 3290 travel time estimates (1-minute medians). Only estimates of which ground truth median was calculated from at least 10 vehicles were accepted to the sample, and 1527 estimates were filtered out.

The tested quality criteria were **Reporting Accuracy** and **Classification correctness**. The method chosen was Reference testing of data collected.

4.5.2. RESULTS

The results from the quality tests are presented in the following Table 10.

Table 10: Results from the Finnish quality tests for Travel Time Information

Criteria	Filtering condition	Value	Quality level
Reporting accuracy (95%)	minimum of 10 vehicles in the ground truth sample	10.3%	* Basic
Reporting accuracy (95%)	minimum of 20 vehicles in the ground truth sample	8.0%	** Enhanced
Classification correctness X% / 20%	minimum of 10 vehicles in the ground truth sample	0.5%	*** Advanced
Classification correctness X% / 10%	minimum of 10 vehicles in the ground truth sample	5.7%	tbd

4.5.3. IMPLICATIONS

4.5.3.1. IMPLICATIONS FOR REPORTING ACCURACY CRITERIA

Due to the travel time measurement algorithm's logic of the tested MAVE-S system, it was necessary to detect also vehicles on the incoming ramp. These vehicles, that consisted 10-30 % of all vehicles in the end of the link, which could not be included in the ground truth sample, and since they may have different speed profile, may affect to the results of the study. Another source for error was the ground truth data system that detected "only" 92% of vehicles, and there may be different detection rate between lanes in congestion. Also the measurement locations differed slightly (~14 meters) due to installation delimitations. Therefore, due to the pilot set-up there were a number of factors causing small measurement deviations between the datasets.

The use of 95 % least accurate measurement as a quality criterion is intended to highlight the importance of production of even quality. For practitioners, it gives a lot of pressure to the study set-up design and the trustworthiness of the ground truth data. Due to this sensitivity for small deviations, the criteria may not be usable for example in the procurement of systems or services where the results have also financial implications.

In reality it may be hard to reach two fully comparable datasets and in such a case it is wise to allow a deviation margin that is not counted for as an error. Change of the criterion to 80 % least accurate measurement would account for this practical need (for the dataset in this study, 80 % percentile would mean 5.2 % deviation, or "error" from the ground truth, close to the quality level ***.

4.5.3.2. IMPLICATIONS FOR CLASSIFICATION CORRECTNESS CRITERIA

The sources for measurement deviations between the two datasets were not affecting the calculation of classification correctness (i.e. error rate), because the criteria allows 20% deviation. Therefore the reliability of the results is very high.

In the current criteria the error threshold remains the same 20% on all quality levels. In practise, many authorities will set targets to higher levels of accuracy and will account smaller deviations than 20 % as errors at least on the highest quality level. Therefore it should be considered whether the error threshold should be tighter on quality levels *** and ****, e.g. 10 %. The classification correctness with 10 % threshold is 5.7 % for the tested dataset.

Even 10 % permitted deviation accounts for the potential deviations due to the test set-up or different samples between the ground truth and the tested data.

4.6. Safety-related traffic information (The Netherlands)

4.6.1. TEST SETTING

Prior to the start of the publication of SRTI-messages, a good method for studying safety related traffic information messages was examined. Assessment methods have been studied within EIP+ for the eight different categories of safety related traffic information messages. One of the methods that would have a good fit for studying the quality of SRTI-messages related to broken down vehicles and accidents is the quality assessment of event based, safety related, traffic information (QSRTI) method. In this, the logging of the time stamps of SRTI-messages is compared to the ground truth. Determining a ground truth, however, currently has very high costs. Therefore, two research methods have been chosen that approach the QSRTI-method and offer insight into the quality of the chain within a short period.

In order to still get an image of how the processes work and what the lead time (latency) of the SRTI-messages is, a case-aimed analysis has been carried out that focuses on the Incident Management notifications (accidents and broken down vehicles). Logging files were extracted from the traffic management centres (UDLS) and incidents feeds (SRTI, ONDA). RTTI data was used to determine the ground truth (speed and flow). In the test the time stamps of messages in the system that sends out the SRTI-messages (TOS) were compared with the time stamps in the logging system in the TMC (UDLS) and local speed data ('ground truth').

In the Dutch quality study several tests were conducted as follows:

1. **SRTI: Unprotected accident area**
   a) Latency/ Reference testing of data collected
   b) Timeliness (update) and Location accuracy / Time-space oriented reference test methods

In the test #1a, logging files were extracted from the traffic management centres (UDLS) and incidents feeds (SRTI, ONDA). RTTI data was used to determine the ground truth (speed and flow). In the test the time stamps of messages in the system that sends out the SRTI-messages (TOS) were compared with the time stamps in the logging system in the TMC (UDLS) and local speed data ('ground truth'). A sample of 100 event messages over a period of 1 year (June 2016 - June 2017) were acquired and the dataset was seen as representative for the whole dataset. Events were randomly selected.

In the test #1b, the logging files were extracted from the traffic management centres (UDLS) and incidents feeds (SRTI, ONDA). RTTI data was used to determine the ground truth (speed and flow). The number of selected cases was 20, all during daytime from the network instrumented with ITS systems. This was necessary to determine the ground truth with enough effects in the traffic flow. The dataset is not representative for the whole dataset, because the determination of the ground truth was only possible for the part of the road network where there is traffic detection every 300-500 meters. An example is given in the figures and table below.

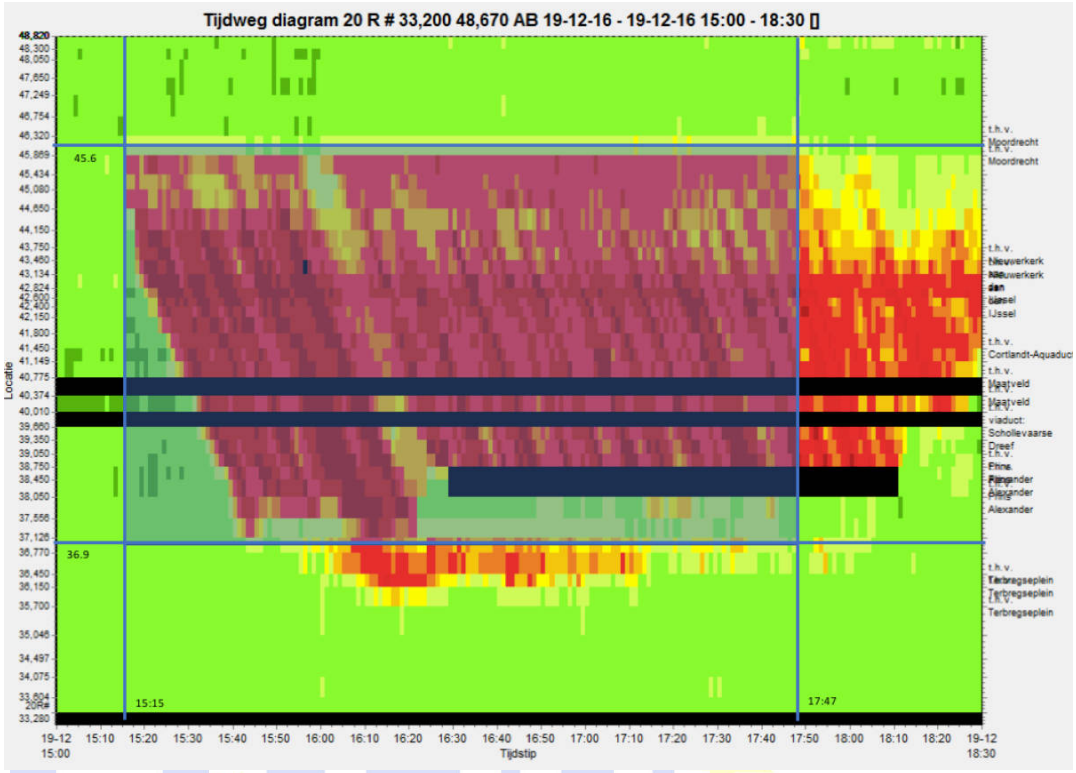


Figure 4: Example of the use of speed contour plot as ground truth in the measurement of latency and timeliness.

Results speed contour plot	Value
Start time	15:15
End time (moment of releasing road)	17:47
Start location	45.6
Maximal length congestion	8,7 km

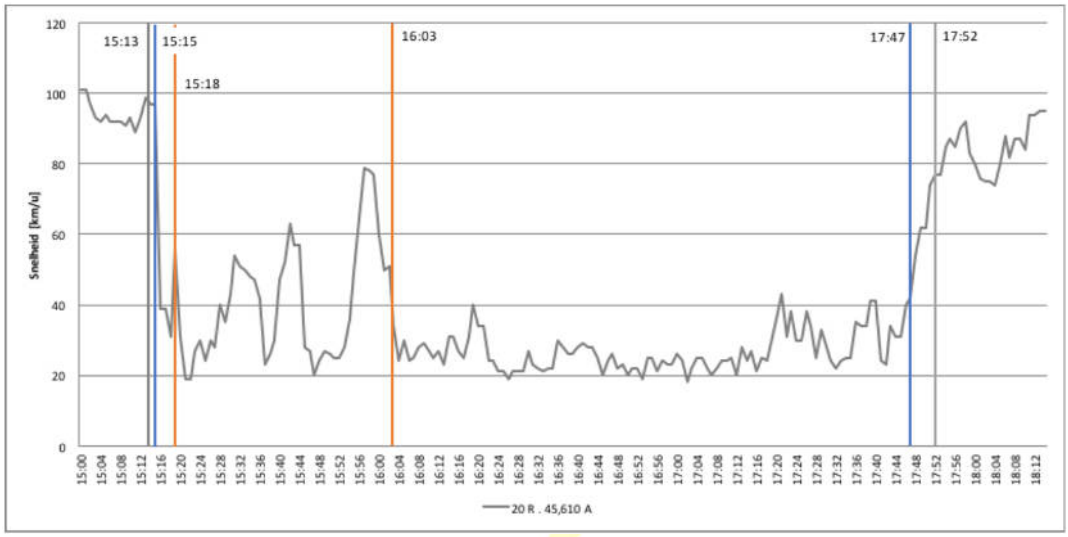


Figure 5: Curve of the speed at the accident including the start and end time of the incident, based on the measured speed (blue), the start and end time logging in UDLS (grey) and the moment of sending and cancelling SRTI-message (orange)

The figure above shows the curve of the speed at the incident, including the start and end time of the incident based on the measured speed (blue), the start and end time logging in UDLS (grey) and the moment of sending and cancelling SRTI-message (orange).

2. SRTI: Animal/people/obstacles/debris on the road (Broken down vehicles)

- Latency/ Reference testing of data collected

In the test #2, logging files were extracted from the traffic management centres (UDLS) and incidents feeds (SRTI, ONDA). RTTI data was used to determine the ground truth (speed and flow). In the test the time stamps of messages in the system that sends out the SRTI-messages (TOS) were compared with the time stamps in the logging system in the TMC (UDLS) and local speed data ('ground truth'). A sample of 100 event messages over a period of 1 year (June 2016 - June 2017) were acquired and the dataset was seen as representative for the whole dataset. Events were randomly selected.

3. SRTI: All types of events

- Classification correctness / Manual verification of events or conditions

In the test #3, comparison of information in logging files extracted from the traffic centres (UDLS) and incidents feeds (SRTI, ONDA) was compared manually. Based on the information in the logging files the true classification of the event was determined again (i.e. ground truth). In case there was not enough logging information, the event was removed from the dataset. Number of cases: ONDA 50, TOS 50 (100 altogether), random sample from a period of 1 year (June 2016 - June 2017).). The dataset (see below) is representative for the whole dataset, events were selected randomly.

Table 11: Number of cases per event category included in the analysis of classification correctness

Category	Amount
Burning vehicle	1
Fuel on the road	1
Defect truck	7
Animal on the road	2
There's debris on the road	3
Cyclist on the road	1
People on the road	2
Mud on the road	1
Obstacles on the road	2
Accident	19
Broken down vehicle	31
Emergency repair	4
Objects on the road	23
Tow-away regulation	2
Reduced visibility	1
Total	100

4. SRTI: Wrong way drivers

- Latency, timeliness end / Reference testing of data collected

In the test #4, timestamps in the logging systems of the TMC are compared with timestamps of the SRTI message system. Also some individual cases (which led to accidents) are described in a qualitative way and related to the incident data. There were 100 event messages over period of 1 year (June 2016 - June 2017), the dataset is representative for the whole dataset, events are selected randomly.

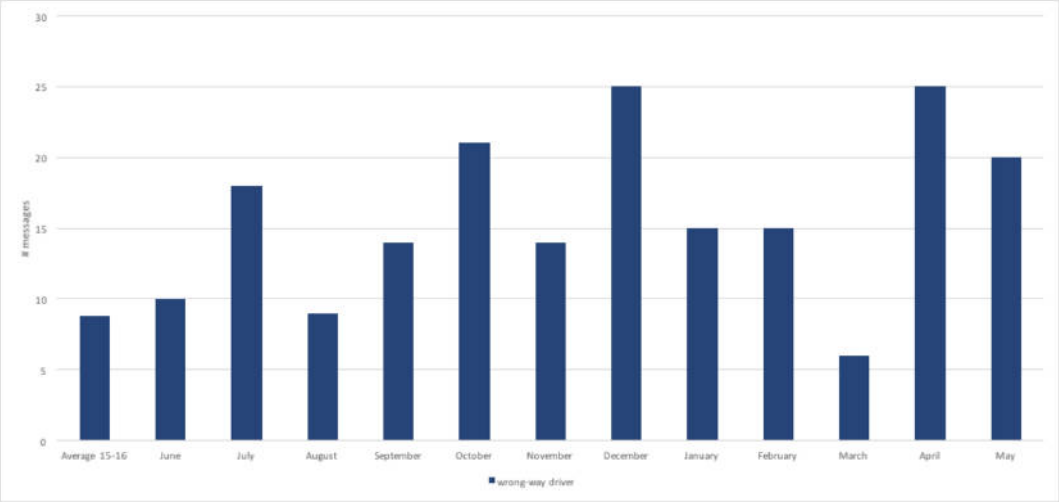


Figure 6: Overview of the total number of SRTI-messages of the category wrong way driver per month during the period from June 2016 to June 2017.

The figure above shows an overview of the total number of SRTI-messages of the category wrong way driver per month during the period from June 2016 up to June 2017.

For a complete report see Hovestad 2017.

4.6.2. RESULTS

The results from the quality tests are presented in the following tables 14-17.

Table 12: Results from the Dutch quality tests for SRTI: unprotected accident area information

Criteria	Specification of test	Value	Quality level
Latency (95%)	test #1a accidents	79% of the messages were sent within 10 minutes from acceptance	below * Basic level
Timeliness start	test #1b		at least *** Advanced
Latency	test #1b		* Basic
Timeliness end	test #1b		at least *** Advanced
Location accuracy	test #1b		at least *** Advanced

Table 13: Results from the Dutch quality tests for SRTI: Animal/people/obstacles/debris on the road

Criteria	Specification of test	Value	Quality level
Latency (95%)	test #2 Broken down vehicles	85% of the messages were sent within 10 mins	below * Basic level

A note from the quality test #2 is that the current definition for latency connects the processing time of IT system to human behaviour. The tight boundaries of the quality levels are therefore not within reach. The reasons for longer latencies are diverse.

Table 14: Results from the Dutch quality tests for SRTI: all events

Criteria	Specification of test	Value	Quality level
Classification correctness	test #3	96% of messages were correctly classified	****

Table 15: Results from the Dutch quality tests for SRTI: Wrong-way drivers

Criteria	Specification of test	Value	Quality level
Latency	test #4	unable to determine.	unable to determine, less than basic level
Timeliness end	test # 4	unable to determine	Timeliness end: advanced level

4.6.3. IMPLICATIONS

Test 1a and 2

The test participant pointed out that the current definition for latency connects the processing time of IT system to human behaviour. The tight boundaries of the quality levels are therefore not within reach in the Dutch conditions. The proposal is to adjust the quality requirement for latency. Another comment is that the systems in the traffic management centre are not developed for this kind of assessment. At this moment the systems consist of different databases and there is no connection between the systems. Therefore the automatisation of the assessment is not currently possible and the required manual work is significant.

Test # 1b

In the Netherlands the location accuracy of the reports is very good, and in that light the requirement levels are generally too low.

Test #3:

Regarding Classification correctness The Dutch test indicates that it is not logical to have exactly the same quality requirement Enhanced and Advanced level.

Regarding classification correctness, the Dutch participant suggests a 2-phase assessment: first check that the event is an SRTI-event (passes criteria for SRTI) and after that check that the event is correctly classified between SRTI categories.

Test #4

Regarding the timeliness update/end, based on the Dutch results it is suggested to tighten the requirement for Enhanced level from Best Effort to 10 minutes.

4.7. Safety-related information for reduced visibility (Spain)

4.7.1. TEST SETTING

The purpose of the study was to analyse the **timeliness start** for a specific case of **SRTI: reduced visibility**. For this case, the timeliness start is conceived as the delay between the occurrence of the event (detected by a visibilimeter) and the first acceptance of the event (considered as the time when the operator signalizes the restriction for a determinate visibility).

For the assessment of this quality criterion the following information was used:

- meteorological data provided by the weather stations located in the area
- signalling record report of the VMS located on the entrance to the problematic stretch of road

It is important to highlight that the clocks of all the devices were synchronised.

For the analysis of the data, it was necessary to process all the information and classify it in order to make it comparable. The information related to meteorological conditions was provided for each 1-minute during the 5 months pilot. This data was grouped considering the visibility restrictions (according to LOS) by using Excel. With the aim of simplifying the study, it was decided to analyse only the information of the most restrictive weather station. The signalling record report included the minute of the restriction change. The difference of time between the visibility level of service (LOS) change and the restriction displayed on the variable message signs (VMS) was calculated for each LOS change situation.

Measurement period of five months (1st January 2017 to 31th May 2017) resulted in a sample of 969 cases corresponding with signalling activations on VMS due to changes on visibility levels. Although the changes in VMS were mainly associated with visibility values, there are some cases -not significant, since represent less than 5 % of the global- that were activated regardless of visibility values (probably because of specific fog episodes with strong winds, rain, high intensity of traffic, etc). Also, some of the information related to February is missing (specifically the last 10 days).

4.7.2. RESULTS

The results from the quality tests are presented in the following table.

Table 16: Results from the Spanish quality tests for SRTI: Reduced visibility.

Criteria	LOS Change	Value	Quality level
GROUP 1: HIGH VISIBILITY TO LESS VISIBILITY			
Timeliness start (95%)	No restriction level to Green Level	18 min	* Basic
	Green Level to Yellow Level	28 min	* Basic
	Yellow Level to Red Level	42 min	* Basic
	Red Level to Black Level	32 min	* Basic
GROUP 2: LESS VISIBILITY TO HIGH VISIBILITY			
Timeliness start (95%)	Yellow Level to Green Level	49 min	* Basic
	Red Level to Yellow Level	38 min	* Basic
	Black Level to Red Level	2 h 22 min	* Basic

If the timeliness start was assessed considering the median value instead of the 95 % quantile, the results would achieve higher quality levels (**enhanced, ***advanced and even **** for the best case).

4.7.3. IMPLICATIONS

Since the LOS changes are activated by TMC operators, the larger sample size does not automatically implicate more accurate results. The variability of the low visibility cases makes it difficult to fix a rigid protocol for the personnel, so the results will always be accompanied by some uncertainty associated with the decision made by the operator. Besides, because of being a manual method, the margin of error is hard to estimate.

From the practitioners' point of view, it is recommended not only to consider the 95 % quantile for the quality level assessment, but to also take into account more indicators like median values and deviations.

5. Evaluation of the results

5.1. Quality criteria for SRTI and RTTI

Most of the quality criteria were tested at least with one quality study, except some of the quality criteria for status type of information (see Table 17).

The test participants gave very few comments regarding the quality criteria itself, i.e. the naming and definitions. A possible reason for this is that the EU EIP 4.1 working group has been further fine-tuning the criteria definitions in the first phases of the project. The validations did not raise many new issues to the discussion.

The Finnish participant indicated, however, that “error rate” is a more descriptive name for the quality criteria for status type of information than "classification correctness". Status type of information is continuous in nature, not categorised, like event type of information is. Furthermore, the common practice is to look at the erroneous side of the measurements.

Therefore, it is recommended to use “error rate” as a quality criteria for status type information in the future. For event type of information, it is recommended to keep using “classification correctness” according to the latest updates in the Quality Package.

Table 17: Quality assessment methods and parts of quality criteria covered by the evaluation studies

	1. Continuous monitoring of equipment performance and availability	2. Manual verification of events or conditions	3. Reference testing of data collected	4. Time-space oriented reference test methods	5. Monitoring of data completeness and latency	6. Regular sampling of message or data content completeness and cor.	7. Verification and calibration of traffic / weather conditions prognosis	8. Surveys of perceived quality by users
Service level parameters								
Geographical coverage								
Availability								
Quality criteria (event)								
Timeliness (start)		ES1	NL1b					
Timeliness (update)		RO1	NL1b, NL4					
Latency (content side)		DK1 RO1	NL1ab NL2, NL4					
Location accuracy		DK1	NL1b DE1-3					
Classification Correctness		NL3, DK1	DE1-3				UK2	
Event coverage			DE1-3					
Quality criteria (status)								
Reporting period								
Timeliness (update)								
Latency (content side)								
Reporting accuracy			UK1, FI1					
Classification correctness			FI1					

Report coverage								
DK1: Assessment carried out according to EIP Practical Guidelines, not according to Quality Package itself								

5.2. Quality requirements for SRTI and RTTI

The implemented quality tests covered most of the quality criteria and requirements. In the following chapters, the implications of the validation tests are discussed and the coverage of each test regarding the quality requirement scale is presented. Each test participant has been given an individual colour, which are explained below each graph. (cf. Figure 7 on p. 42)

Before we start to discuss the implications of the performed tests, it is useful to resemble, what is the reason for the original setting of the quality requirements for RTTI and SRTI, and what is their status in the implementation of the ITS directive in the Member States. Currently the quality criteria are the so called “initial target values” that are presented for stakeholders for building experience and validation (Lohoff et al. 2015). The validation work started in EIP+ project and is not further continued in EU EIP. Hence, it should be noted that the tested requirements are still initial values and do not yet carry more official status.

In the design of the initial target values it was noted that because circumstances and priorities across Europe differ, it is not possible to find a suitable single value for a certain service type that would be applicable in all Member States. Also it is necessary that the quality requirements accommodate differences in the implementation of ITS systems. Therefore, the following logic was set for the quality framework structure (Lohoff et. al. 2015):

- the basic level (minimum requirement) should be achievable in all Member States
- the levels should accommodate European diversity
- further levels should act as guidance for quality improvement in the future

In this task, the results from the quality tests are evaluated against the above mentioned design principles of the quality framework. In practise, there are two implications of this:

- if there are more than one country/stakeholder that fail to reach the minimum requirement, it is an indication that the initial value may have been set too high and the requirement should be lowered. If there is only one country/stakeholder that has implemented the test and failed, lowering of the requirement can be considered if the development level of ITS systems is of high quality in the country.
- if more than one countries/stakeholders reach the quality level Advanced (or in particular level ****), it is an indication that the initial value may have been set too low to leave incentives for future development, and the requirement should be tightened. Tightening of the requirement can be considered, if there is only one test participant.

Regarding the fourth quality level, it was decided in the EUEIP 4.1 working group, that we have still too little knowledge of the possible quality achievable with future or near-future technologies, such as G5. It is likely, that setting values that follow a trendline from quality level 1 onwards based on current technology, as done in the current initial values, will lead to too low quality requirement for the fourth quality level. Therefore, it was

decided to keep the fourth level in the quality framework to show that there is improvement potential after the advanced level, but leave the requirement itself blank for all criteria.

5.2.1. QUALITY REQUIREMENTS FOR RTTI EVENT INFORMATION

Four participants performed validation studies for RTTI event information. The coverage of the validation studies regarding these services is presented in the following graph.

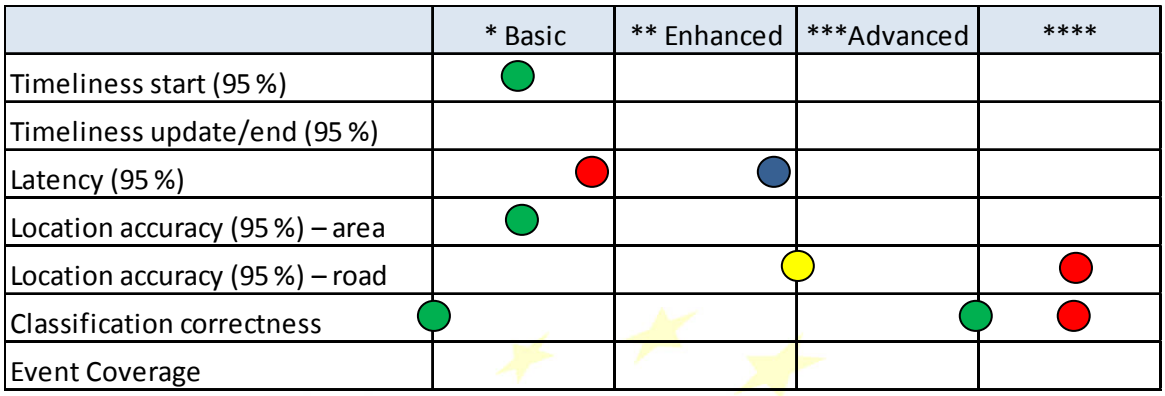


Figure 7: Coverage of validation studies regarding RTTI event information (Real-time event information / Traffic conditions information (TCI) / Weather information). Red bullets for the Danish tests, green bullets for tests performed in the UK, the dark blue for tests performed in Romania and yellow for the German test.

All tests except one gave results that passed the minimum quality requirement. The **classification correctness** in the UKs test fell below the minimum requirement (15%) for Ice event and snow events, whereas in the same test the classification correctness for rain events was in fact at advanced level. As the test in question handled road weather forecasts, and as the prediction of snow events is known to be particularly difficult, the test result does not generate the need to change the initial quality requirements.

The test regarding **timeliness start** indicated that for some services such as weather services the delays may be still substantial. This favours the current minimum requirement “Best Effort”.

The Danish test for **latency** provided result that passed the minimum quality requirement clearly, as there were only 3 events where latency was slightly over 5 minutes out of 53 total events. Also the Romanian test showed very low latencies for the limited event sample.

The only point that raises to discussion is the fact that for Classification correctness the requirement is the same (10 %) for both Enhanced and advanced levels. There were two studies that reached the advanced level, another one actually level ****. It is recommended to tighten the requirement for levels *** - **** , since there is no theoretical nor practical grounds for allowing up to 10 % erroneous content on the most advanced quality levels.

The German test was focused on **location accuracy** for road work warnings along roads. Looking at the accuracy expressed as a geographic deviation, it was concluded

that Advanced *** level could be fulfilled easily. This resulted in a recommendation to tighten the requirements for the Enhanced ** and Advanced *** levels, taking into account that road work information must be somewhat more accurate than the initial 5 or 10 kilometres requirements. However, the German test failed for the Basic level *, which requires a reported event to be within the "right link between Intersections". This was due to georeferencing issues between different datasets, which are related to the applied assessment method. To address this dilemma in future evaluations, it is recommended to apply a 1 km marginal (500 meters to both direction from the centre point of an intersection) in the definition of an "intersection". In addition, because intersections allow the car drivers to change route in case of an incident, it was seen crucial from users' perspective that the event is in the correct link between intersections also on higher quality levels, even though the actual distance based requirement is also applied. Therefore from Enhanced level forward, the data must pass two parallel quality requirements.

5.2.2. QUALITY REQUIREMENTS FOR RTTI STATUS ORIENTED INFORMATION

There were two participants performing quality tests for status oriented information, namely for travel time information. The tests covered only Reporting accuracy and Classification correctness parameters. The coverage of the validation studies regarding the quality requirement scale is presented in the following graph.

	* Basic	** Enhanced	***Advanced	****
Reporting period (95 %)				
Timeliness update (95 %)				
Latency (95 %)				
Reporting accuracy (95 %)	●	●		
Classification correctness			●	
Report Coverage				

Figure 8: Coverage of validation studies regarding RTTI status oriented information: Travel time information. Green bullet for the test performed in UK and blue for tests performed in Finland.

Regarding **reporting accuracy**, both tests passed the minimum requirement, as the results were on both sides of the threshold between Basic and Enhanced levels. This does not directly indicate a need to change the requirements. However, both performed studies raise the 95% percentile into discussion. Based on the test setting descriptions of these two studies, it seems that the primary source for measurement deviations for travel time information may many times be the ground truth data collection systems, which are rarely bullet-proof in smaller scale technical pilots or separate quality tests. Similar experiences were found both in the Finnish and UK tests for travel time information. The main concern is the fact that the quality requirement for reporting accuracy targets to the 95% percentile of the measurement data distribution. This is the part of the distribution where it may well be that the measurement difference is explained by the properties of the ground truth data, not the studied data itself. This remark makes the use of the

criteria as such (95% least accurate measurement) somewhat speculative and limits it use towards internal quality assurance processes, unless the percentile in the requirements is changed.

From the practitioners' point of view, the **classification correctness** criterion works much better in many real-life use cases. It allows certain amount of measurement deviation, which is inherent in processes comparing data from different sources, and concentrates on measurements that exceed the set threshold. The Finnish validation study exceeded the limit for Advanced level so clearly, that it is an indication that the threshold for the higher quality levels may be too low. In fact, in the current criteria the error threshold remains the same 20 % on all quality levels, but in practise, many authorities will set targets to higher levels of accuracy. Therefore it should be considered to tighten the error threshold on quality levels *** and ****, e.g. to 10 %.

5.2.3. QUALITY REQUIREMENTS FOR SRTI: WRONG WAY DRIVER

The Dutch participant validated two quality requirements for SRTI: Wrong way driver, namely timeliness update/end and latency. The coverage of the validation studies regarding these services is presented in the following graph.

	* Basic	** Enhanced	***Advanced	****
Timeliness start (95 %)				
Timeliness update/end (95 %)			●	
Latency (95 %)	●			
Location accuracy (95 %) – area				
Location accuracy (95 %) – road				
Classification correctness				
Event Coverage				

Figure 9: Coverage of validation studies regarding SRTI: Wrong way driver. Tests performed in the Netherlands.

For **timeliness update/end**, the result of the test was that exactly 95% of the events had timeliness of 5 minutes or less, hence the quality reached level *** Advanced. The report discussed whether this indicates that the threshold for the advanced level is too low. However, as the Dutch test is the only validation test for the service in question, and as the threshold is already much tighter than for the RTTI events (10 minutes), it is not recommended to change the initial value for the advanced level. However, it should be discussed whether there should be a definite requirement for the enhanced level (such as less than 10 minutes).

For **latency**, the test result was that for 83% of events the latency was less than the minimum requirement (<3 minutes). Therefore the quality is below the minimum requirement level. It should be noted that for SRTI: wrong way driver the latency requirements are stricter than for other SRTI events, for which the threshold is <10 minutes. This result is clearly an indication that the initial values, especially for the

minimum requirement, may be too strict for most of the Member States at the time being, since the development level of ITS is high in the Netherlands.

In Denmark the traffic information policy in case of wrong way drivers is, that the TMC operator first concentrates his/her efforts on broadcasting the warning via radio. This prioritization is based on the fact that at the moment radio reaches more drivers in a short period of time, which is the need in case of such incidents. Since the radio broadcasting takes time a couple of minutes, it was indicated by the Danish partner that being able to distribute the information via NAP in three minutes is too strict a requirement to be fulfilled in most events. This was acknowledged in the revision of the minimum requirement for latency for wrong way drivers.

5.2.4. QUALITY REQUIREMENTS FOR ALL SRTI EVENTS/CONDITIONS EXCEPT WRONG WAY DRIVER

A fair amount of validation activities concerned the other SRTI events than wrong way driver. The coverage of the validation studies regarding these services is presented in the following graph. Three participants provided results to these services.

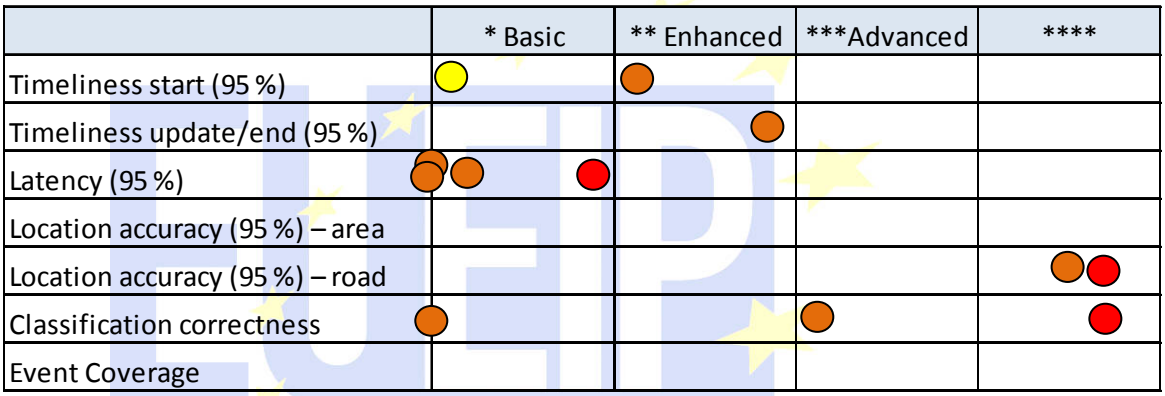


Figure 10: Coverage of validation studies regarding SRTI: all events except wrong way driver. Orange tests performed in the Netherlands, yellow in Spain and red in Denmark.

For **timeliness start** the Spanish study presented rather long timeliness values in the 95% percentile of the performance range regarding harsh weather warnings, even though the median values were on a good level. This indicates that for the time being allowing Best Effort as the minimum requirement is justified, if the 95% percentile is kept as focus in the requirement. In the Dutch study the timeliness values for unprotected accident areas and broken down vehicles events were manually studied for a limited sample. The results indicated that in 95% of cases timeliness value of 9 minutes or less was reached, hence the quality was on level ** Enhanced.

For **timeliness end** the same Dutch study gave a result that in 82% of events the timeliness end value was less than 10 minutes, hence the *** Advanced was not reached. However as the minimum requirement is Best effort, and as the median timeliness value was small (less than 2 minutes), the result does not give grounds for lowering the requirements.

For **latency**, both Dutch tests (test for accidents and broken down vehicles) indicated that the current quality level for latency is below the basic level, i.e. for more than 5% of

the cases the latency is longer than 10 minutes. In those tests 79%-85% of the messages reached the minimum requirement level 10 minutes. The test report indicated that because the current definition for latency connects the processing time of ITS systems to human behaviour, the tight boundaries of the quality levels are not within reach at the moment. Similar results were found in the Finnish quality test for RTTI: incident information in the EIP+ project, where the 95% percentile for Latency was over 23 minutes. In all of these cases the average latency is on a very acceptable level around 2-3 minutes, but the problem is that there is not yet an even quality in the process, and in around 20% of cases the latency grows over 10 minutes or even much longer.

However, there were two studies where the minimum requirement was reached. In the manual Dutch study of 20 events, the latency just reached the minimum quality level with 1 event out of 20 having latency over 10 minutes. In the Danish study for accidents the result was that for only 3 events out of 53 latency was over 5 minutes, a result that almost passed the requirement for level ** Enhanced. The Danish result applies for both RTTI events and SRTI events (unsecured accident scene).

The conclusion regarding latency is, that the current requirement is likely to be beyond reach in many Member States, regarding the fact that the Netherlands in one of the most developed countries in traffic information services. Proposal for altering the requirement is presented in the next chapter.

There was one study for **location accuracy**, provided by the Dutch participant using manual verification of a small sample of 20 events. The result was that for 95% of events the accuracy was better than 1 km measured on the road. There was only 1 event out of 20 where the accuracy was between 1-3 km. The result passes the quality requirement for level ****. Taking into account the fact that the initial values on all levels, especially on levels Enhanced and Advanced are rather low from the viewpoint of the road users (even advanced level allows deviations below 5 km), it is recommended to tighten the requirements thorough all levels of the quality requirement. Similar conclusion was drawn from the German quality test regarding RTTI events.

There was one study looking at **classification correctness**. The study prepared by the Dutch participant tested the classification correctness produced by the TMC (TOS) and also an automatic logging system in place (UDLS). Both samples contained 80 randomly selected events. The result was that 90.2% of messages sent out by the TMC were correctly classified as “objects on the road”. This passes the quality requirement of the Advanced level. The performance of the logging system was somewhat different, only 45% of events were correctly classified. The Dutch report indicated that it is necessary to improve the assessment whether or not an event is SRTI event. The conclusion of the test is that the minimum requirement level for classification correctness is on a reasonable level, but it is advisable to tighten the requirements for higher quality level *** Advanced and ****. There are no specific grounds for keeping the requirement the same for Enhanced and Advanced levels. Similar conclusion was drawn from the Finnish study regarding RTTI status information.

It should be mentioned that also the Danish study for accidents included the location accuracy and classification correctness criteria, and the results apply for both RTTI and SRTI services.

5.3. Quality assessment methods

5.3.1. RESULTS ON QUALITY ASSESSMENT METHODS IN EVALUATION STUDIES

The quality assessment methods and RTTI and SRTI services covered by evaluation studies are summarised in Table 18.

All except one study used either *Manual verification of events or conditions* or *Reference testing of data collected* as a research method. This coverage of the methods toolbox reveals which methods are the most practical ones to be used in such discontinuous quality measurements, and therefore should not be seen as a critical deficiency in this validation round.

The partners performing evaluation studies were requested to provide their positive and negative findings related to the quality assessment methods documented in the EIP Quality Package. The reported positive and negative findings are summarised in

Table 19.

Table 18: Quality assessment methods and RTTI and SRTI services covered by the evaluation studies

	1. Continuous monitoring of equipment performance and availability	2. Manual verification of events or conditions	3. Reference testing of data collected	4. Time-space oriented reference test methods	5. Monitoring of data completeness and latency	6. Regular sampling of message or data content completeness and correctness	7. Verification and calibration of traffic / weather conditions prognosis	8. Surveys of perceived quality by users
RTTI - Event information		DK1	DE1,3					
RTTI - Traffic condition information								
RTTI - Travel time information			UK1 FI1					
RTTI- Weather information		RO1					UK2	
SRTI - Temporary slippery road								
SRTI - Animal/people/obstacles/debris on the road		NL3	NL2					
SRTI - Unprotected accident area		DK1 NL3	NL1ab					
SRTI - Short term road works			DE2					
SRTI - Reduced visibility		ES1						
SRTI - Ghost driver (or wrong-way driver)			NL4					
SRTI - Unmanaged blockage of a road		NL3						
SRTI - Exceptional weather conditions								
DK1: Assessment carried out according to EIP Practical Guidelines, not according to Quality Package itself								

Table 19: Summary of positive and negative findings related to Quality assessment methods and parts of quality criteria covered by the evaluation studies

Summary of positive and negative findings reported in quality assessments		
Quality assessment method	**Positive findings**	**Negative findings**
2. Manual verification of events and conditions	+ The method was easy to apply with Excel, there was no need for special software (ES1)	– No negative findings, the only difficult thing related to the study was data collection (ES1)
3. Reference testing of data collected	+ It was easy to calculate the results with simple spreadsheet methods (FI1) + Probe data collection procedures are in place. Results are easy to calculate by using Excel. (UK1) + The method gives insight into the process and time span of the event through the visual presentation of the ground truth data (NL2) + Reference testing can be supported by external (third-party) data sources, considering the trend that much of SRTI/RTTI-related reality can be detected by parties outside the traffic management centers nowadays. In particular, FCD data from service providers is a powerful source to detect current traffic status in a (timely and geographically) complete manner.	– There are some known difficulties in calculating accurate link journey times from GPS data for links that are very short or that run very close to other roads (UK1) – Due to financial constraints, sample sizes are low, but this method has been suitable to Transport Scotland to date. (UK1) – No negative findings, challenges were related only to the study set-up and the comparability of data sets (FI1) – Determination of the ground truth is time consuming and it is not possible to select the events randomly. To determine ground truth, we used cases that occurred during a certain period of the day and at the part of road network where we have a high density of ITS. So the result of the sample is not representative. (NL2) - When the ground truth is based on an external (third-party) data source, there are technical challenges when bringing together different "data worlds". In this case, the usage of FCD-based event information from a private service provider raised some questions on consistent georeferencing and a common interpretation of "what is an individual event". Further, the "ground truth assumption" for an external data source is not automatically proven. (DE)
7. Verification and calibration of traffic / weather conditions prognosis	+ Easy to calculate metrics using simple spreadsheet methods. (UK2)	– Event identification and weather station data extraction was a manual process, as this was the first exercise of its kind. If this was to be carried out more regularly we would write a database script and parametrise the queries. (UK2)

The templates filled for each of the evaluation studies were also analysed to identify any improvements proposed to the quality assessment methods. These were found in field

"proposals for improvements of the assessment method" of the template as well as in the field "Proposals for improvements of the EIP documentation (incl. Quality criteria)". The identified proposals for improvements to the quality assessment methods are summarised in Table 20.

Table 20: Summary of improvements proposed to quality assessment methods

Summary of improvements proposed to quality assessment methods	
Quality assessment method	**Proposed improvements**
Unidentified	- Install systems for automatic extraction of the required quality data from the IT systems processing the traffic information messages (DK1)
2. Manual verification of events and conditions	- Not only consider the 95th percentile for assigning the quality level. More indicators such as median and deviation should be taken into account. (ES1)
3. Reference testing of data collected	- Note should be added of the use of travel time (speed) median. High share of trucks of the traffic flow may cause large deviations in study results, if different vehicles are detected by the system providing the ground truth and the tested system. For example, the median of travel time may be affected by the composition of the traffic flow. (FI1) - Add recommendation of the sample size of the calculated travel time medians in ground truth (20) and also a recommendation of well-designed study set-up minimising systematic errors in measurement (FI1) - The document (Quality Package) does not cover aspects of data cleansing and basic quality control measures that can be continually monitoring and ensure ongoing quality parameters for travel time data. Equipment installation, set-up, static link and journey time route configuration also contribute to the quality of journey time system data. (UK1) -When referencing to external (third-party) data sources, the consistency and comparability of the assessed data has to be ensured. Further, the "ground truth assumption" has to proven BEFORE the actual quality assessment, including validation (and potentially calibration) of the external data source. (DE)
7. Verification and calibration of traffic / weather conditions prognosis	- Description of how the metrics (described with the quality assessment method) tie into the quality tables (described in the Quality Package) (UK2) - Description of the method refers to (Wilks 2005) which is no longer available, and a new edition has been published. (UK2) - Heidke skill score can be simplified from the version presented in the document. This simplified version would be more practical for calculation. (UK2)

In addition to clearly positive and negative findings and proposal for improvement, also other results and conclusions on quality assessment methods were reported during the evaluation studies. These other results and conclusions are summarised in Table 21.

Table 21: Summary of other results and conclusions related to quality assessment methods

Quality assessment	Summary of other results and conclusions related to quality assessment methods
UK2	There is no mapping between the metrics presented as a part of the description of quality assessment method 7 (Verification and calibration of traffic /weather conditions prognosis) and the elements of the quality criteria presented in the quality package (e.g. the relationship between "proportion correct" and "classification correctness).
NL1, NL3 and NL5	Systems in the traffic management centre are not developed for this kind of assessments. At the moment, the systems consist of different databases and there is no connection between the systems. Automatisation of the quality assessment is not possible in the current situation, and this complicates the process of getting results.

5.3.2. DISCUSSION ON INDIVIDUAL QUALITY ASSESSMENT METHODS

Reference testing of data collected

The collection of the data set to be used as ground truth may require manual work. For example, this was the case in one of the quality assessments carried out in the Netherlands (NL2). In the first of the quality assessments in UK (UK1), the sample size was limited by financial constraints for the work. In conclusion, the sample size available for analysis may be limited.

The constraints for sample size – due to the rare occurrence of events, financial constraints or other reasons – should be taken into account in the quality criteria and the descriptions of the quality assessment methods when possible. While it may be difficult to make major changes to the quality criteria, it is possible to update the description of quality assessment method "Reference testing of data collected" with a guidelines how to deal with low sample sizes. This guideline has to be provided separately for nominal variables (e.g. LOS or road weather class) and continuous variables (e.g. travel time, timeliness and latency).

In case of reference testing of data collected, challenges may also be related to minimising the impact of systematic errors when collecting data and comparing different data sets (quality assessments UK1,FI1 and DE). The quality assessment methods have been described on a general level in the EIP Quality Package to ensure that the descriptions are applicable to different elements of the quality criteria and to different RTTI and SRTI services. On the other hand, very specific advice may be required for applying the quality assessment method with individual types of data sets and data collection methods. This is the case for example with difficulties calculating travel times based on GPS positions of vehicles (UK1) as well as detecting road works with FCD data from a service provider (DE). Challenges, that are very specific to certain data collection method or quality criterion being analysed, can be addressed by providing examples of related quality assessments.

The description of the quality assessment method "Reference testing of data collected" does not include an advice how data with continuous values (e.g. travel time) shall be aggregated over time when comparing it to ground truth. For example, the impact of

random variation on the median of travel time may be large if the time slot being analysed includes only few observed travel time values. The need to aggregate individual observed values over time arises when a data set with individual observations is compared to aggregated data set. The question on aggregation of data over time has been encountered e.g. with travel times (FI1).

Note should be made of the use of travel time (speed) median. Let's assume that a travel speed median is calculated from 11 vehicles. If the sample consists of 6 personal cars and 5 trucks, the expected value in free flow conditions in Finland is 120 km/h, but if the sample consists of 6 trucks and 5 personal vehicles, the expected value is 90 km/h. At night time with small samples and high share of trucks potentially causes large deviations in the study results, if partly different vehicles are detected by the ground truth and the tested system. Therefore it is recommended to use large enough sample size in the ground truth data, for travel time info a suitable limit could be at least 20 vehicles.

Manual verification of events and conditions

Manual verification of events and conditions can be used with different RTTI and SRTI services and with different quality criterion. Manual verification is usually not possible for a large number of events or conditions. This means that the sample size collected for analysis will usually be limited, and it is not always possible to achieve statistically significant results. This should be made clear in the description of the quality assessment method.

Verification and calibration of traffic / weather conditions prognosis

This quality assessment method may require either manual work with analysis of the data or implementing a dedicated software (UK2). The description of the method includes several metrics taken from literature, but they have expressed separately from the quality criteria. It would be preferable to connect the metrics included in the description of the method to the quality criteria for RTTI and SRTI services described earlier in the Quality Package. This can be achieved by creating a mapping between parts of the quality criteria and the metrics described in the quality assessment method. In addition, some minor points for improvement were identified during quality assessment (UK2) such as replacing the reference to (Wilks 2005) with a reference to the latest edition and a simpler description of the Heidke skill score.

6. Proposal for the validated quality recommendation for SRTI and RTTI

6.1. Quality criteria for SRTI and RTTI

There was only one proposal for changing the quality criteria, namely to replace Classification correctness with Error Rate for status oriented information. Error rate is more descriptive criteria for continuous information types that are not to be categorised such as event type of information.

Having different quality criteria for event and status type of information adds the number of criteria in the framework, but the fact is that most of the criteria are already different for the different service types, it can be argued that this change would not increase the complexity as much as it would increase the understandability of the framework for practitioners.

The proposed changes are highlighted with red ink to ease the reading of the report.

Table 22: Proposed change to the quality criteria based on evaluation studies.

Definition of Criteria for RTTI and SRTI		event information	status oriented information
Classification correctness	100% - percentage of the published events which are known to be not correct (concerning actual occurrence of this event type / class). *(NB: It is recommended to use time and place for matching events between data sets.)*	X (Yes)	~~X (Yes)~~ (No)
Error Rate	The percentage "A" of the reported locations for which the relative inaccuracy of the status report is larger than percentage "B" in a given reporting period. Expressed as A/B. - ~~redefinition of "classification correctness" for status reporting~~	(No)	X (Yes)

6.2. Quality requirements for SRTI and RTTI

The proposed changes to the initial target values are presented in the following tables with red ink to ease the reading.

6.2.1. QUALITY REQUIREMENTS FOR RTTI EVENT INFORMATION

Table 23: Proposed changes to the quality requirements for RTTI Event information

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start ~~(95%)~~	**Best effort**[1]	95 % of events Acceptance[2] after first detection < 10 min[1] Time between occurrence and first detection: Best effort[1]	95 % of events Detection & acceptance[2] < 5 min after event occurrence	~~Detection & acceptance[2] < 3 min after event occurrence~~
Timeliness update/end ~~(95%)~~	**Best effort**[1]	Best effort[1]	Detection & acceptance[2] 95 % of events < 10 min after event change/end	~~Detection & acceptance[2] < 5 min after event change/end~~
Latency (content side) ~~(95%)~~[3]	~~**< 10 min**~~[1] **80% of events < 10 min**[1]	~~< 5 min~~[1] 80% of events < 5 min[1]	~~< 3 min~~ 95% of events < 5 min[1]	~~*< 2 min*~~
Location accuracy ~~(95%)~~ - Area	95% of events **administrative region (TCI: n/a)**	Geographic area; 95 % of events 10 km accuracy *(TCI: n/a)*	Geographic area; 95 % of events 5 km accuracy *(TCI: n/a)*	~~*Geographic area; 2 km accuracy (TCI: n/a)*~~
Location accuracy ~~(95%)~~ - Road	**95 % of events correct link between Intersections**[4,5]	**95 % of events correct link between Intersections**[4,5] AND distance < ~~10~~ 4 km[4]	**95 % of events correct link between Intersections**[4,5] AND distance < ~~5~~ 2 km[4]	~~*< 2 km*~~[4]
Classification correctness	**> 85%**	> 90%	~~> 90%~~ > 95%	~~*< 2%(TCI: < 5%)*~~
Event coverage	**Best effort**	Best effort[6]	> 80% of all occurring events	~~*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an organisation's quality policy*

[3] *Counts for the presented share of the analysed data address the latency measurements at the start and at the end of events.*

[4] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.*

[5] *A marginal of 500 meters to both directions from the geographical centre point of the intersection is applied to address the differences in georeferencing systems. Ramps leading to rest areas only are not defined as intersections.*

[6] *Definition of criterion and value need further examination.*

6.2.2. QUALITY REQUIREMENTS FOR RTTI STATUS ORIENTED INFORMATION

Table 24: Proposed changes to the quality requirements for RTTI Status information

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Reporting period ~~(latency)~~	**Best effort**	5 min	1 min	
Timeliness update ~~(95%)~~	**Best effort**	95 % of reports < 5 min	95 % of reports < 2 min	
Latency (content side) ~~(95%)~~	**95 % of reports < 10 min**	95 % of reports < 5 min	95 % of reports < 2 min	~~*< 1 min*~~
Reporting accuracy ~~(95%)~~	**80 % of reports < 20%**	80 % of reports < 10%	80 % of reports < 5%	
~~Classification correctness~~ Error rate	~~**90%/80%**~~ **10%/20%**	~~95%/80%~~ 5%/20%	~~98%/80%~~ 5%/10%	~~98%/90%~~
Event coverage	**Best effort**	>90%	>97%	

EU EIP European ITS Platform

6.2.3. QUALITY REQUIREMENTS FOR SRTI: WRONG WAY DRIVER

Table 25: Proposed changes to the quality requirements for SRTI: Wrong way driver

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start ~~(95%)~~	**Best effort**[1]	95 % of events Acceptance[2] after first detection < 5 min[1] Time between occurrence and first detection: Best effort[1]	95 % of events Detection & acceptance[2] < 2 min after event occurrence	~~Detection & acceptance[2] < 1 min after event occurrence~~
Timeliness update/end ~~(95%)~~	**Best effort**[1]	Best effort[1]	95 % of events Detection & acceptance[2] < 5 min after event change/end	~~Detection & acceptance[2] < 2 min after event change/end~~
Latency (content side) ~~(95%)~~[3]	~~**< 3 min**~~[1] **80% of events < 5 min**[1]	~~< 2 min~~[1] 80% of events < 2 min[1]	~~< 1 min~~ 95% of events < 2 min	~~*< 1 min*~~
Location accuracy ~~(95%)~~ - Area	**95% of events Administrative region (TCI: n/a)**	Geographic area; 95% of events 10 km accuracy *(TCI: n/a)*	Geographic area; 95% of events 5 km accuracy *(TCI: n/a)*	~~*Geographic area; 2 km accuracy (TCI: n/a)*~~
Location accuracy ~~(95%)~~ - Road	**95% of events affected road segment not longer than 50 km or road number with destinations (city names) nearby**[4]	95% of events correct link between Intersections[4,5]	95% of events correct link between Intersections[4,5]	~~*< 5 km*~~[4]
Classification correctness	**> 75%**	> 85%	> 90%	~~>90%~~
Event coverage	**Best effort**	Best effort[6]	> 80% of all occurring events	~~*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an*

organisation's quality policy

[3] Counts for the presented share of the analysed data address the latency measurements at the start and at the end of events.

[4] For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.

[5] A marginal of 500 meters to both directions from the geographical centre point of the intersection is applied to address the differences in georeferencing systems. Ramps leading to rest areas only are not defined as intersections.

[6] Definition of criterion and value need further examination.

6.2.4. QUALITY REQUIREMENTS FOR ALL SRTI EVENTS/CONDITIONS EXCEPT WRONG WAY DRIVER

Table 26: Proposed changes to the quality requirements for SRTI: all events except Wrong way driver

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Timeliness start ~~(95%)~~	**Best effort**[1]	95% of events Acceptance[2] after first detection < 10 min[1] Time between occurrence and first detection: Best effort[1]	95% of events Detection & acceptance[2] < 5 min after event occurrence	~~Detection & acceptance[2] < 3 min after event occurrence~~
Timeliness update/end ~~(95%)~~	**Best effort**[1]	Best effort[1]	95% of events detection & acceptance[2] < 10 min after event change/end	~~Detection & acceptance[2] < 5 min after event change/end~~
Latency (content side) ~~(95%)~~[3]	~~**< 10 min**~~[1] **80% of events < 10 min**[1]	~~< 5 min~~[1] 80% of events < 5 min[1]	~~< 3 min~~ 95% of events < 5 min[1]	~~*< 2 min*~~
Location accuracy ~~(95%)~~ - Area	**95% of events Administrative region**	~~Administrative region~~ Geographic area; 95 % of events 10 km accuracy (TCI: n/a)	Geographic area; 95% of events ~~10~~ 5 km accuracy	~~*Geographic area; 2 km accuracy*~~
Location accuracy ~~(95%)~~ - Road	**95% of events correct link between Intersections**[4,5]	95% of events correct link between Intersections[4,5] AND distance < ~~10~~ 4 km[4]	95% of events correct link between Intersections[4,5] AND distance < ~~5~~ 2 km[4]	~~*< 2 km*~~[4]
Classification correctness	**> 85%**	> 90%	~~> 90%~~ > 95%	~~*< 2%*~~ ~~*(TCI: < 5%)*~~
Event coverage	**Best effort**	Best effort[6]	> 80% of all occurring events	*> 90% 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] *An event is considered accepted when it has been found trustworthy according to an organisation's quality policy*

[3] *Counts for the presented share of the analysed data address the latency measurements at the start and at the end of events.*

[4] *For an event with a length, like many road works, both the start point and the end point must fulfil the given criteria.*

[5] *A marginal of 500 meters to both directions from the geographical centre point of the intersection is applied to address the differences in georeferencing systems. Ramps leading to rest areas only are not defined as intersections.*

[6] *Definition of criterion and value need further examination.*

6.3. Quality assessment methods

The recommendations for improvements to the descriptions of the quality assessment methods are described in the table below.

Table 27: Proposed improvements to quality assessment method descriptions in the Quality Package

Quality assessment method	Proposed improvements
Reference testing of data collected	A guideline on how to deal with low sample sizes should be added. This guideline has to be provided separately for nominal variables (e.g. LOS or road weather class) and continuous variables (e.g. travel time, timeliness and latency).
Reference testing of data collected	In case of challenges, that are very specific to certain data collection method or quality criterion being analysed, easy to understand and well documented examples should be provided on applying the quality assessment method.
Reference testing of data collected	Guidelines on aggregating data over time periods should be provided at least on general level. For example, the sample size should be large enough in the ground truth data. In case of travel time, a suitable limit could be at least 20 vehicles.
Manual verification of events and conditions	The description of the quality assessment method shall be updated with guidance how to deal with low sample sizes
Verification and calibration of traffic / weather conditions prognosis	A mapping shall be created between the parts of the quality criteria and the metrics described in the quality assessment method.
Verification and calibration of traffic / weather conditions prognosis	The reference to (Wilks 2005) shall be replaced with a reference to a newer edition of the same document, if the content of the new edition matches the older one.
Verification and calibration of traffic / weather conditions prognosis	A simpler description of the Heidke skill score shall be included in the description of the method.

6.4. Practical examples and interpretation of the requirements

One conclusion from the validation efforts was that especially for practitioners that are not familiar with the Quality Framework some of the quality criteria and requirements may be difficult to fully understand, or at least a there are risks for misinterpretation. To ease the understanding and improve the applicability, it was decided to add practical examples to the Quality Package. The practical examples should be introduced in the package straight after the introduction of the quality criteria and requirements.

Two models of such practical examples are presented below. Practical examples for all quality criteria should be introduced in the next version of the Quality Package.

Example1. Measuring Latency for SRTI events&conditions except wrong way driver

In the Netherlands Rijkswaterstaat measured latency and timeliness end for SRTI service regarding wrong way driver. In the test the time stamps of messages in the system that sends out the SRTI-messages (TOS: Datex Converter, system for logging and pushing of all SRTI categories, except accidents and broken down vehicles) were compared with the time stamps in the logging system in the TMC (UDLS: incident registration system; traffic managers log their events and actions here). A sample of 100 wrong way driver messages were acquired. The quality test proceeded as follows:

For Latency

- sort the events in spreadsheet in the order of growing latency (see table 1).
- check what is the cumulative percentile that passes the threshold for minimum level requirement (< 3 min)
- only 86 % of events have latency smaller than 3 minutes, hence the tested dataset does not pass the original minimum quality requirement (95 % of events < 3 min). The Basic level isn't reached.
- testing against the proposed new requirement for latency, it can be seen that the dataset passes the minimum requirement (80% of evens are <5min) for basic level and even passes the requirement for enhanced level (only 80% of events are <2min). Enhanced level is reached.

Category	Amount	Proportion	cumulative
0 minutes	62	62%	62%
0-1 minutes	11	11%	73%
1-2 minutes	10	10%	83%
2-3 minutes	3	3%	86%
3-4 minutes	7	7%	93%
4-5 minutes	0	0%	93%
5-6 minutes	2	2%	95%
6-7 minutes	3	3%	98%
7-8 minutes	1	1%	99%
8-9 minutes	1	1%	100%
>9 minutes	0	0%	100%

Distribution of latencies in the dataset

Example1. Measuring Latency for SRTI events&conditions except wrong way driver

For Timeliness End

- sort the events in spreadsheet in the order of growing timeliness end (see table 2).
- Old and new quality requirements for Timeliness End are quite similar. With 95% the data set just passes the minimum requirements for the advanced level. With 95% the minimum level of 95%. For Timeliness End the Advanced level is reached.

Category	Amount	Proportion	cumulative
0 minutes	81	81%	81%
0-1 minutes	10	10%	91%
1-2 minutes	2	2%	93%
2-3 minutes	1	1%	94%
3-4 minutes	0	0%	94%
4-5 minutes	1	1%	95%
5-6 minutes	1	1%	96%
6-7 minutes	0	0%	96%
7-8 minutes	0	0%	96%
8-9 minutes	1	1%	97%
>9 minutes	3	3%	100%

Distribution of Timeliness End in the dataset

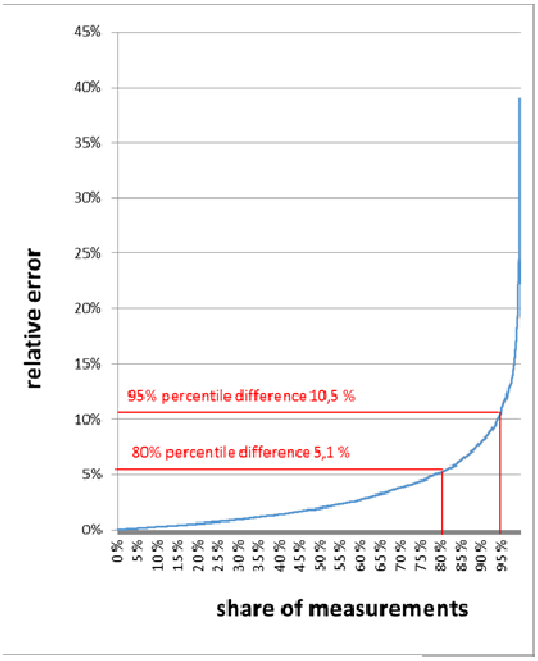
Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Original requirement for latency (95%)	<3min	<2min	<1min	<1min
Proposed new requirement for latency	80% of events < 5 min	80% of events < 2 min	95% of events < 5mins	-
Original requirement for Timeliness Update/End (95%)	Best effort	Best effort	detection and acceptance <5min after event change/end	detection and acceptance <2min after event change/end
Proposed new requirement for Timeliness Update/End	Best effort	Best effort	95% of events detection and acceptance <5min after event change/end	-

Quality requirements for Latency and Timeliness End (Original and proposed)

Example2. Measuring reporting accuracy for RTTI Status type of information (Travel time information)

In Finland the Finnish Transport Agency tested the accuracy of connected induction loop measurement system in the measurement of travel times. In the test data was collected from a 2 km long road link, and it was compared to ground truth data collected with LPRS system. Measurement period of 4 working days resulted in a sample of 3290 travel time estimates (1-minute medians). The quality test proceeded as follows:

- remove estimates that are based on ground truth sample of less than 10 vehicles (1527)
- calculate measurement difference (=error) between the tested data and ground truth for each 1-minute interval
- calculate the absolute values of the errors, and then calculate the relative error (%)
- sort the whole dataset in the order of growing relative error
- check from the data what is the relative error that corresponds to the 95th percentile in the error distribution
- Since the 95th percentile relative error is 10,5%, the data passes clearly the requirement for the minimum level (20%), but remains slightly below the threshold for the enhanced level (10%)
- Testing against the proposed new requirements, it is checked from the error distribution that the 80th percentile of errors is 5,1%.
- It is concluded that the data passes clearly the new proposed requirement for the enhanced level (80% below 10%), but misses slightly the requirement for the advanced level (80% below 5%).



Distribution of errors in the dataset

Parameter	* (Basic)	** (Enhanced)	*** (Advanced)	****
Orinigal requirement for Reporting accuracy (95%)	20 %	10 %	5 %	
Proposed new requirement for Reporting accuracy	80 % of reports <20%	80 % of reports <10%	80 % of reports <5%	

Quality requirements for reporting accuracy

7. Conclusions

In this task of EU EIP sub-activity 4.1 ("Task 3.0 - Quality Test for RTTI and SRTI") another round of validations was carried out, this time with better understanding of the challenges regarding such studies among the active partners, which were mostly the same as in the EIP+ quality test task. There were seven participants in the study and they performed a total of 13 quality studies, many of them covering several quality criteria. All services were covered by at least one study except RTTI Traffic condition information and SRTI Temporary slippery road and Exceptional weather conditions. For event-type of services all quality criteria were validated by at least one study, even though separate requirements may be set for different service types. On this round none of the reported studies yielded fully non-conclusive results. From this point of view it can be said that the validation round has been successful.

The most important implication from the validation studies was that the working group recommends changes to some of the definitions for quality criteria and quality requirements. These recommendations will be discussed with EU EIP sub-activity 4.1 partners and other stakeholders, and will potentially lead to a future revision of the RTTI & SRTI Quality Package.

For RTTI and SRTI event information services, the recommendation is to change the requirements for **latency**, **location accuracy (Road**) and **classification correctness**. For **latency** the validation tests proved that the initial values were set so high that many countries would be likely to fail to reach the minimum level, and therefore the requirement was slightly lowered. On the other hand for **location accuracy** and **classification correctness the** requirements could be tightened to increase ambition for future development**.**

For RTTI status information the validation studies only considered **reporting accuracy** and **classification correctness**. The implication was that for **reporting accuracy** the criteria should exclude 20% of the far end of the measurement data distribution (instead of only 5%), due to the fact that there tends to be some amount of measurement deviation inherent in study settings including travel time ground truth data. For **classification correctness** the recommendation was to tighten the requirements on the two higher quality levels. Finally it was recommended to change the criterion **classification correctness** for status type of information back to error rate, which is more descriptive criteria for such continuous data types.

Regarding the fourth quality level, it was decided, that we have still too little knowledge of the possible quality achievable with future technologies such as G5. It is likely, that setting requirement values that follow a trendline from quality level 1 onwards, as done in the initial values, will lead to too low quality requirement for the fourth quality level. Therefore, it was decided to keep the fourth level in the quality framework to show that there is improvement potential after the advanced level, but leave the requirement itself blank for all criteria.

To sum up, the implemented validation tests were of high quality and yielded better understanding of the quality of traffic information services in Europe as well as appropriate quality requirements for such services. Taking into account the coverage of the successful validation studies, the updated quality requirements can be given a status of “validated target values” instead of “initial target values”.

From theoretical perspective it should be noted, however, that many of the requirements were validated only by one study, performed in one member state, with a limited data sample. Moreover, one study can only validate one quality level at a time, leaving the other parts of the quality requirement scale untouched. Therefore, it is recommended to continue the collection of quality test results from various member states in the coming years, and to perform qualitative analysis of such a compilation. If such an analysis proves that the quality requirements are not on a correct level a process should be agreed how to alter the quality requirements. In addition, it is recommended to fulfil the fourth quality level requirements, when enough evidence is collected from certain future technologies and their impact on service quality.

Regarding quality assessment methods, all except one study used either *Manual verification of events or conditions* or *Reference testing of data collected* as a research method. This lesser coverage of the methods toolbox reveals which methods are the most practical ones to be used in such discontinuous quality measurements, and therefore should not be seen as a critical deficiency in this validation round. The tests yielded several proposals how to improve the method descriptions so that they would best serve the future users.

To assist a data supplier such as a Traffic Information Centre (TIC) or a Traffic Management Centre (TMC), who wants to determine the quality of their SRTI (Safety Related Traffic Information) and RTTI (Real Time Traffic Information) event messages (not including “Wrong way driver”), a relatively short guide has been written: “Quality of Safety Related and Real-time Traffic Information, Practical Guidelines”. These practical guidelines have been tested in one of the performed tests with good results. But more tests will be needed for a better evaluation of these guidelines.

It is recommended to update the RTTI & SRTI Quality package and the Practical Guidelines according to the proposal in this report.

8. References

Hovestad Marlous, S. 2017. Self-assessment SRTI. Study to the quality of safety related traffic information messages during the period from June 2016 up to June 2017. Rijkswaterstaat Netherlands.

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.

Kulmala, R., Öörni, R., Laine, T., Lubrich, P., Schirokoff, A., Hendriks, T. Quality package for safety related and real-time traffic information services. Deliverable of EIP+ 3.1 Testing and Validating of the Quality Recommendations and Results for Traffic Information from EIP. Version 1.07 delivered within the EUEIP Activity 4.1. on 10th November 2017.

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.

Rystrøm, L. 2017. Quality of Safety Related and Real-time Traffic Information, Practical Guidelines, version 0.9, 8 December 2017.