Handbook on Management of data quality for train performance analysis

RailNetEurope
Austria Campus 3
Jakov-Lind-Straße 5
AT-1020 Vienna

Phone: +43 1 907 62 72 00
Fax: +43 1 907 62 72 90
mailbox@rne.eu
www.rne.eu

Content

Document History

Table 1: Document history

Version	Author	Date	Changes
00.1	Simona Di Loreto	02/10/2013	First draft
01.0	RNE	04/10/2013	Revised by RNE JO
01.1	RNE	14/10/2013	Text added on page 4 according to SBB) ´s (Karl Guntern) proposal
01.2	RNE	04/11/2013	Comments received before the OAS meeting on 14-11-2013 (AP, SN,PvK) + additional items by RNE IT staff
02.0	RNE	12/11/2013	Modification according to decision of the OAS on November 14th 2013
02.1	RNE	13/01/2014	Revision before the meeting
3	WG	21/01/2014	Revision during the meeting
4	RNE	13/02/2014	Final draft
4.1	RNE	24/03/2014	Comments by sub-group
4.2	RNE	28/07/2014	Minor modifications; update of annexes and appendixes; agreed by the DQCs in the meeting of July 24th 2014
5	RNE	10/02/2015	Overall revision
5.1	RNE & DQCs	06/03/2015	Modifications after remarks and DQCs meeting (25/02/2015)
5.2	RNE & DQCs	28/03/2015	Approved version with fine tuning after written remarks by DQCs. Fine-tuning of the glossary
5.3	RNE & DQCs	28/05/2015	Revision during the DQCs meeting
6.0	RNE	23/10/2017	Complete revision of Handbook including introduction to new setup and TIS DQ Management procedures
7.0	RNE	06/04/2022	Complete revision
8.0	RNE	13/10/2022	Document Revision and Annexes preparation
8.1	RNE & DQEG	17/10/2022	Annexes Revision and fine-tuning during DQEG meeting

Abbreviations Used in this Handbook

Table 2: Abbreviations used in Handbook

Abbreviation	Explanation
CTT	Contracted Timetable
DQI	Data Quality Indicator
DQ	Data Quality
DQE	Data Quality Expert
DQH	Data Quality Handbook = Handbook on Management of data quality for train performance analysis
DWH	Data Warehouse
ERA	European Railway Agency
KPI	Key Performance Indicator
IM	Infrastructure Manager
OAS	Oracle Analytic Server – RNE tool for reporting and analysis
PAX	Passenger
RA	Running Advice
RFC	Rail Freight Corridor
RNE	Rail Net Europe
RNE MB	RNE Managing Board
RNE GA	RNE General Assembly
RU	Railway Undertaking
TIS	Train Information System
TIS CCB	Change Control Board
TIS TB	TIS Technical Board
TPM	Train Performance Management

Foreword

TIS has become the European widely recognised international tool for real-time monitoring. This and the permanent demand for reports from RFCs, RUs, IMs, Terminals, European Commission, ERA and others leads to higher requirements on quality and range of data and thus the need for a systematic approach to permanent data quality improvement.

On 5 May 2020, the RNE General Assembly agreed on the **Data Quality Strategy for reporting,** stating several missions and objectives to fulfil the vision to:

"Reach the level of data quality to become reliable and respected international reports provider and European train running data warehouse, based on commonly agreed principles and procedures without extra manual efforts with clear topology, descriptions and rules"

To fulfil the objectives stated within the Data Quality Strategy for reporting, 3 follow-up data quality projects were started:

- **Project 1: Basic requirements on data sources for reporting purposes**
  - to define and agree on the standard requirements on data delivery and processing
- **Project 2: Data Quality monitoring and management**
  - to ensure the implementation and fulfilment of the basic requirements on data sources
- **Project 3: Basic requirements for RFC TPM reporting**
  - to define and agree on the common standards to be applied for RFC TPM reports

As a result of Project 1, the '*RNE Guidelines on Basic TIS requirements on data delivery for reporting purposes'* (further referred to as Guidelines) were approved by the RNE General Assembly on 19 May 2021. These Guidelines provide the description of basic requirements within the existing technical and functional frame of TIS on data delivery and standard TIS procedures and processes regarding train linking. The Guidelines serve as the basic reference for setting up the data quality monitoring process.

Moreover, during the RNE General Assembly on 19 May 2021, members agreed on RNE TIS Declaration (further referred to as Declaration), and specifically agreed on:

- sending complete and correct data to TIS in accordance with methods and rules defined in the *RNE Guideline "Basic TIS requirements on data delivery for reporting purposes"* and in compliance with the TAF/TAP TSI;
- participating and actively cooperating in monitoring and improvements of TIS data quality following the agreed processes described in this *Handbook on Management of data quality for train performance analysis* (further referred to as Data Quality Handbook);

To describe the processes of regular monitoring and improvement of data quality and to define basic data quality indicators monitoring in compliance with the above-mentioned Guidelines and Declaration, this Data Quality Handbook was prepared within Project 2 and approved by the RNE General Assembly in May 2022.

This Data Quality Handbook covers the following topics:

- Definition of scope of data quality management for train performance analysis
- Actors involved in data quality management and their tasks and responsibilities
- Indicators used to measure data quality and related data quality reports
- Processes and procedures covered by data quality management
- Tools supporting the management of data quality

Scope of RNE Data Quality Handbook

The framework proposed by this document is aimed to provide a structured procedure to regularly check, analyse and improve the quality of the data provided by data providers and processed by TIS to ensure the reliable performance monitoring.

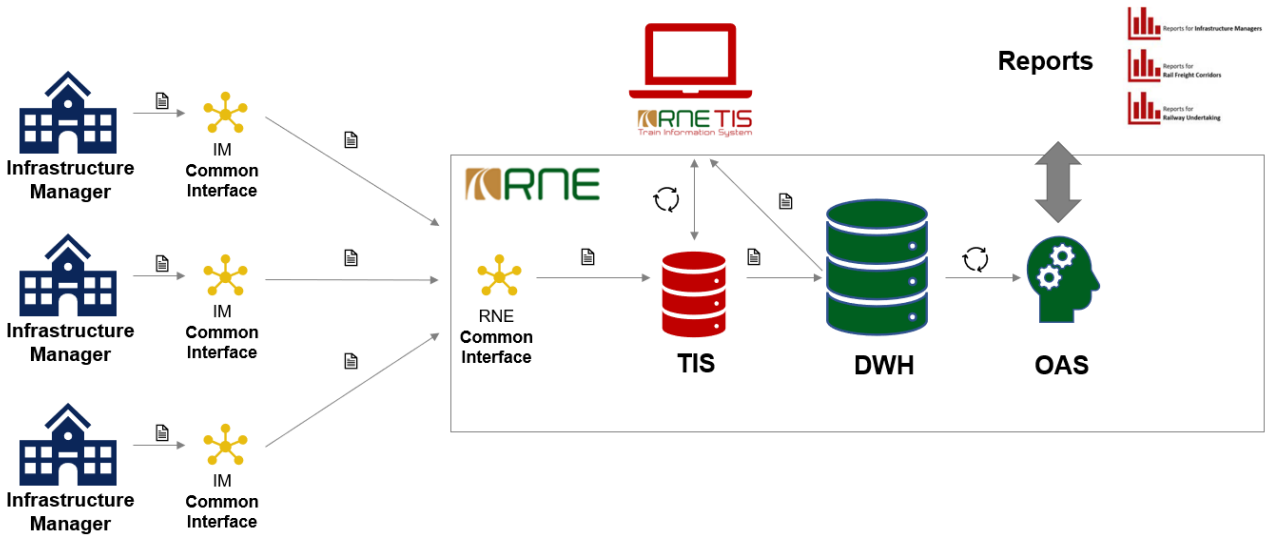


Figure 1 Process from IM messages to complex performance reports

To ensure a reliable performance monitoring, technical quality and data quality must be on sufficient level. Therefore, a close cooperation between IT experts, data quality experts and performance experts is crucial. Detailed cooperation rules and responsibilities of these key players is described in the next chapters.

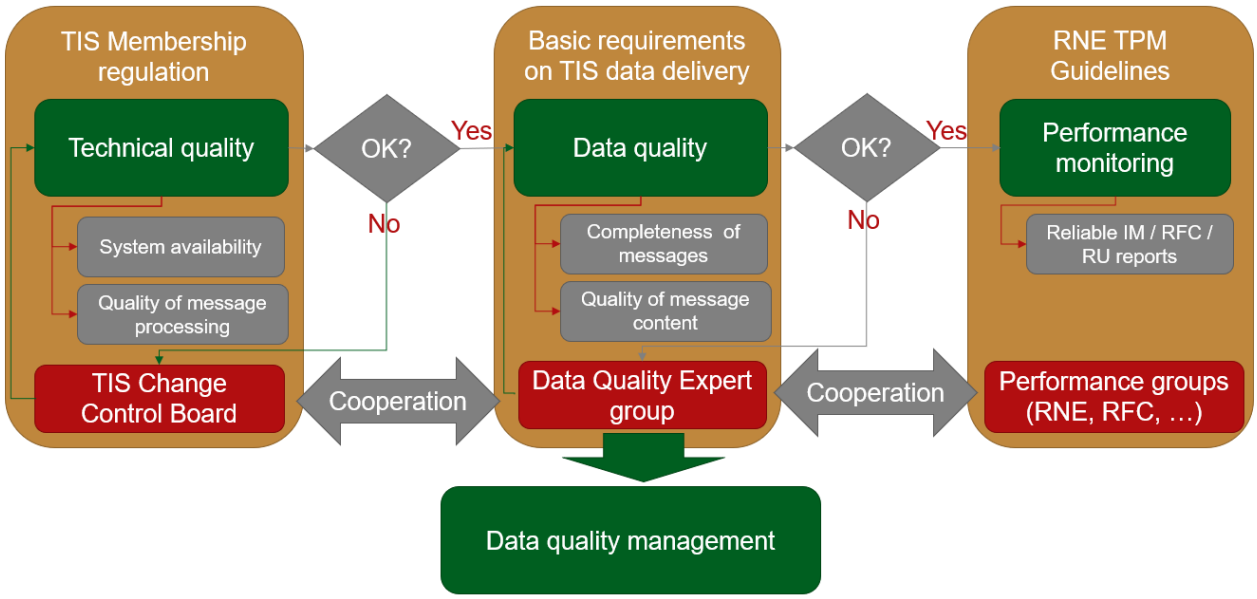


Figure 22 Process cooperation between stakeholders

Technical Quality

TIS is a system managed by RNE, therefore the main responsibility for guaranteeing that TIS provides high data quality relies on RNE. In this context, RNE must take care that the system is always running in an efficient and correct way.

To do so, RNE performs checks and keeps contact with the IT supplier and monitors the system availability via Service Level Agreements (SLAs) with all data providers. Monitoring of system availability and quality of message processing is covered by TIS Membership regulation and handled on the level of TIS Change Control Board.

The TIS Change Control Board is also deciding about any changes and further developments of TIS functionalities. Therefore, in case new developments or changes are needed in order to improve e.g. data quality, and thus also reports' reliability, the TIS Change Control Board has to be addressed.

Data Quality for Reporting Purposes

While technical quality can focus on system availability and technical correctness of messages, additional checks are needed to confirm, that all required messages are sent to TIS and provide all required information to ensure reliable performance monitoring.

When data sent by an IM is not complete or correct, on the one hand, RNE has limited possibilities to find out the cause(s) and on the other hand, most of the times, RNE is not able to solve problems internally caused by the IM's national systems or procedures.

Therefore, both TIS users and data providers (IMs) must be deeply involved in the procedures for checking and improving the data quality. Main focus of these procedures is not on TIS availability, and technical correctness of the messages delivered to TIS, but on the quality of information provided within messages send to TIS, its processing by internal TIS and DWH procedures and its completeness and quality for the needs of precise train performance management purposes.

The basis for data quality measurement are the requirements described in the RNE Guidelines Basic TIS requirements on data delivery for reporting purposes, focusing on:

- **Basic requirements on data delivery:**
  - Messages to be delivered
  - Trains to be reported within messages
  - Locations to be reported within messages
- **Train linking procedures:**
  - To ensure the proper linking of national train run section into single international train run
- **Border section information**
  - To collect all mandatory information about each border section to ensure:
    - Clear identification of trains crossing the border
    - Reliable performance reports on border section level, including dwell time measurement
    - Precise monitoring of the data quality, e.g. linked vs. unlinked trains

The process how to monitor, analyse and improve the compliance with the above-mentioned guidelines is the main scope of data quality management and is described in detail in this Data Quality Handbook.

Performance Monitoring

Several groups are regularly using or intend to use RNE performance reports based on the TIS data, e.g.:

- RFC Performance management working groups
- RFC task forces
- Quality Circles Operations
- Bilateral groups (IM-IM, Cross-border)
- European entities (ERA, DG Move)

The management of development and reliability of performance reports is handled on RNE Performance Management Working group level. This group also plays a role of Reporting Change Control Board, deciding not only about the reports, but also about a suitable RNE Reporting tool.

Different performance management working groups, while working with performance reports, are strongly depending on reliability of the reports and are usually the first ones that notice inconsistencies within data quality. Therefore, regular cooperation and information exchange of performance experts with data quality experts is crucial. This cooperation and communication are also described within this Data Quality Handbook.

Commitment Monitoring

As experiences during previous years have shown, the main success key to improving data quality is a high and regular commitment of all data providers. Therefore, monitoring of data providers' commitment, as stated in RNE TIS Declaration, is also covered by this handbook.

Strong commitment is required:

- on expert level, to properly analyse the current status and propose actions required for quality improvement;
- on management level, to allocate necessary resources and budget to implement actions required for quality improvement.

Monitoring of commitment will be regularly performed and reported to RNE and RFC Managing level, and will cover monitoring of:

- participation and active involvement of IM's representatives in data quality processes covered by this Data Quality Handbook
- ability to reach agreed data quality targets and perform agreed actions to improve the data quality.

The precise methodology of monitoring the commitment of all involved parties is included in Annex 1.

External Influences on Data Quality

Operational behaviour of external partners, (e.g., the way in which RUs request a path) and incompatible operational rules and national laws might have a significant impact on data quality as well. This influence on data quality is not always solvable by the IMs themselves in short term. However, these issues shall be identified and addressed to the concerned parties.

Data Quality Management Process

The basic data quality management process is defined by 3 main activities:

- Monitoring of the data quality status – identification of data quality problems
- Analysis of reasons influencing data quality
- Identification and implementation of corrective actions

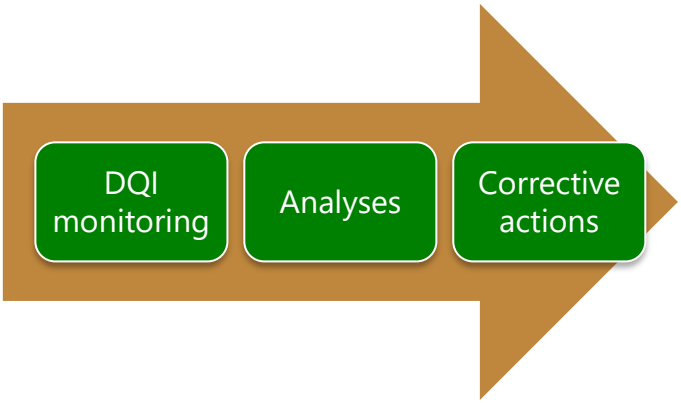


Figure 3: Basic scheme for Data Quality Management Procedure

To ensure this deep and regular cooperation and cover all aspects related to data quality, several DQ levels are introduced:

- RNE DQ Expert level
- RFC DQ Expert level
- RNE DQ Management level
- RFC DQ Management level

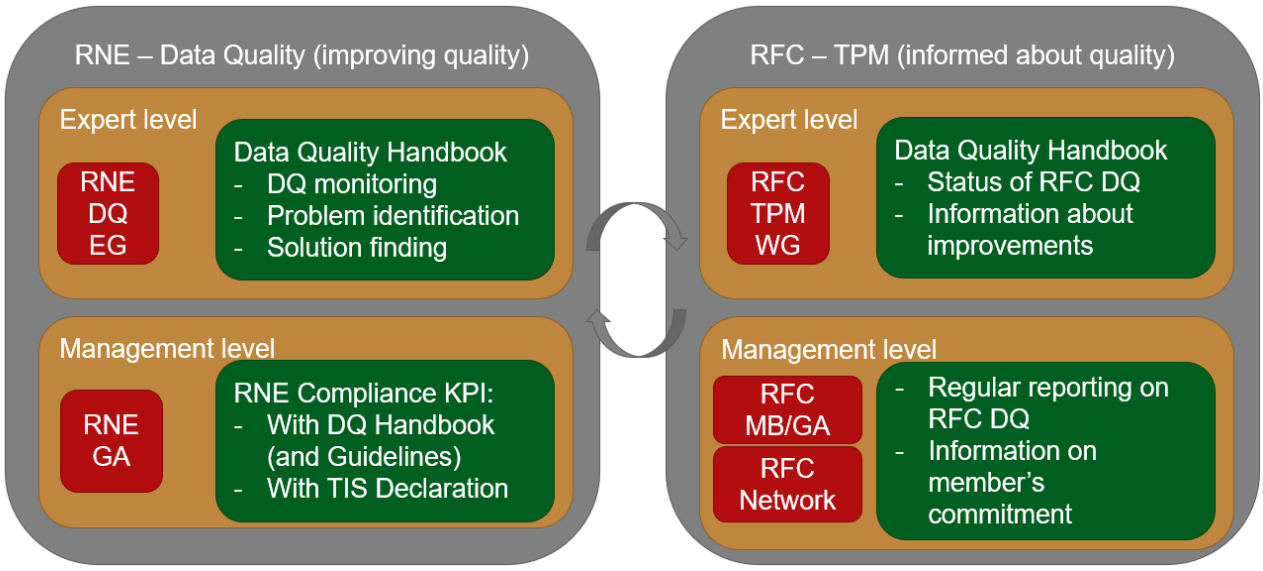


Figure 4: Data Quality Levels

The focus of the RNE data quality management processes is on the complete international traffic, both freight and passenger. For RFC-related processes, the focus stays only on RFC-related, international freight traffic.

RNE Data Quality Expert level

The goal of RNE expert level is to analyse the data quality on the most detailed level to find root problems and define appropriate measures to eliminate problems.

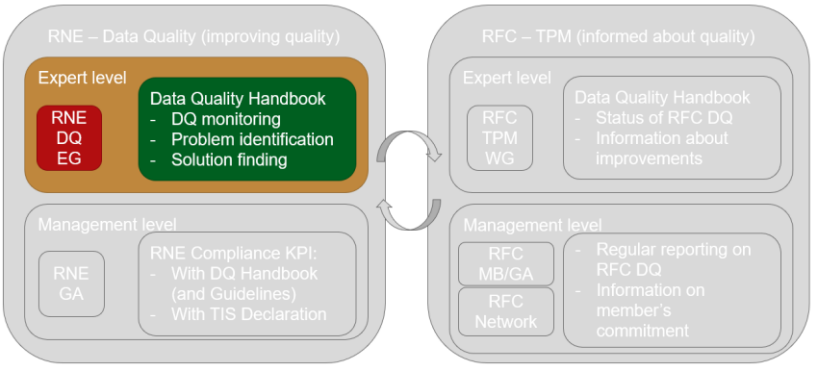


Figure 5: RNE Data Quality Expert level

The basic data quality management process is performed:

- on bilateral level between relevant data provider (IM) and RNE, focusing on the specific IM data
- within RNE Data Quality Expert group, focusing on overall data quality

The main tool for data quality monitoring and improvement on RNE Data Quality Expert level includes:

- Basic data quality management procedure
- Half-year evaluation of data quality – Data Quality Targets and Actions
- Regular discussion on RNE Data Quality Expert group level

Basic data quality management procedure

A data quality problem could be caused by IT problems, by a not suitable procedure in planning or operation phases, therefore involvement of all related national company departments is needed. In these cases, each IM must grant that both analysis and corrective actions are performed in due time and efficiently. At this aim, each IM nominates a “single supervising contact person” for TIS data quality issue - Data Quality Expert (DQE) - whose tasks will be:

» Being the contact partner for RNE for all data quality issues related to train performance analysis
» Actively participate in the RNE Data Quality Expert group and expert level activities
» Being regularly involved and informed about the status of fulfilment of data quality targets
» Responsible for forwarding requests of action/clarification to internal responsible persons
» Monitoring the internal problem-solving process and reporting to RNE its status
» Serving as the main contact person for data quality related questions also to other data quality levels and partners

Data quality management procedure is applied monthly. Once the process proofs to go smoother (no reporting problem, communication flows up and running, updates procedures fully working) it can be replaced by a quarterly procedure. The aim of the procedure is to keep data quality under constant control.

The procedure consists of the following steps:

- RNE ensures loading of the data at the beginning of each month regarding previous month's train runs
- Data quality report is executed and automatically delivered to Data Quality Experts by 7th of each month for the previous month
- If more detailed analyses are needed, data quality expert can execute additional reports directly in RNE Reporting tool (OAS) to target the problem more efficiently
- All partners check the report and react to urgent problems
- RNE and/or Data Quality Expert populate the list of shortcomings, based on the data quality reports created
- All partners check the list of shortcomings and provide feedback on problems indicated by 20th of each month
- Regular meetings are organised where data quality and reporting problems are discussed

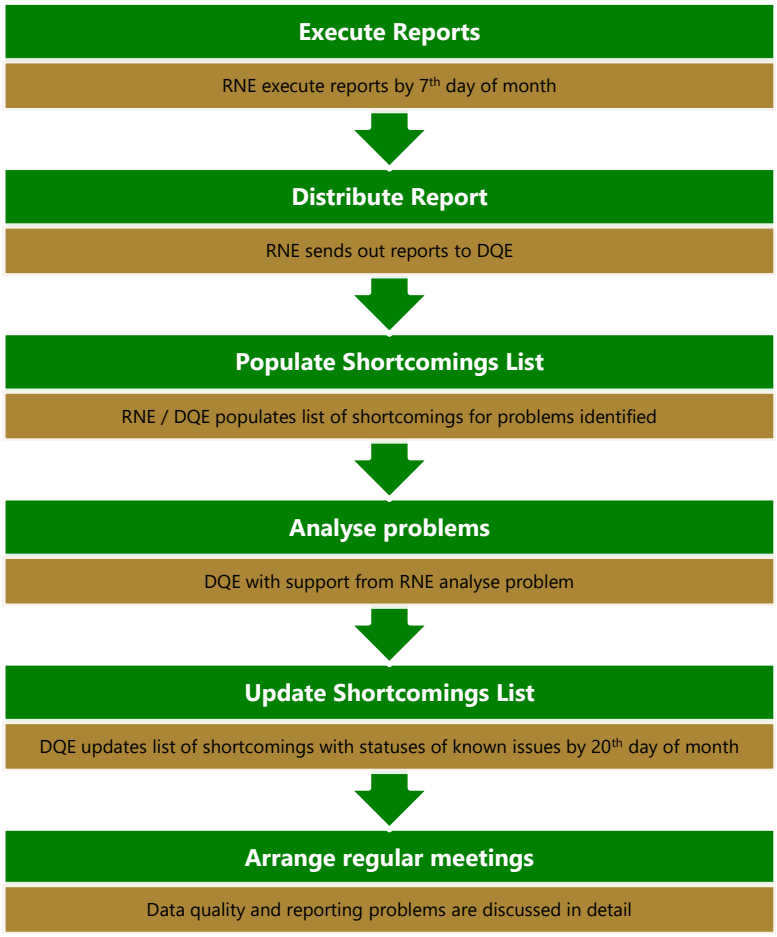


Figure 6: Data Quality check process

Data Quality Experts have direct access to the RNE Reporting tool (OAS), where they can execute and download reports specifying different parameters as wanted.

More detailed analyses can be done upon request by RNE.

The outcome of the problem identification and analysis shall be a list of problems the so called "shortcoming list". Handling will be done by an online web-based tool. Every Data Quality Expert is obligated to use this tool and update status of any known issue listed. Additionally, he/she can populate the List of shortcomings with new issues.

More details about the list of Shortcomings can be found in chapter 0.

Steering of data quality improvement – Data Quality Targets and Actions

To steer the data quality process on RNE / IM expert level, specific data quality targets and actions to reach these targets are agreed between RNE and every single IM providing data to TIS.

Each IM shall nominate a Data Quality Expert (DQE) and a Data Quality High-Level Representative (DQ HLR), with the authority to allocate resources within its IM to work on identified and agreed actions. If no special Data Quality High-Level Representative is nominated, the RNE GA representative shall take over this responsibility. The list of nominated experts and high-level representatives is maintained by RNE in the Data Quality Management tool.

To monitor, whether the basic requirements stated in RNE Guidelines *'Basic TIS requirements on data delivery for reporting purposes'*, are fulfilled by each single IM, a set of standard data quality indicators was defined and is described in chapter 0. These indicators are calculated for each IM and an IM specific target value to be reached within a defined period is mutually agreed with RNE, taking into account the current status of data quality and also the duration of TIS membership.

IM Data Quality Expert together with RNE expert will evaluate the current data quality status and jointly propose possible target values, improvements and actions to be done and specify also resources needed to perform the proposed actions. Main source of information to prepare the action list is the list of shortcomings. Proposed targets and actions are provided to the Data Quality High-Level Representative for evaluation, agreement and commitment to allocating necessary resources and budget. This process of proposing and approving targets and actions is handled and completely supported by the Data Quality Management tool.

By approving targets and actions in the Data Quality Management tool the Data Quality High-Level Representative agrees with target values and commits to providing sufficient resources and efforts to achieve these targets and perform the proposed tasks.

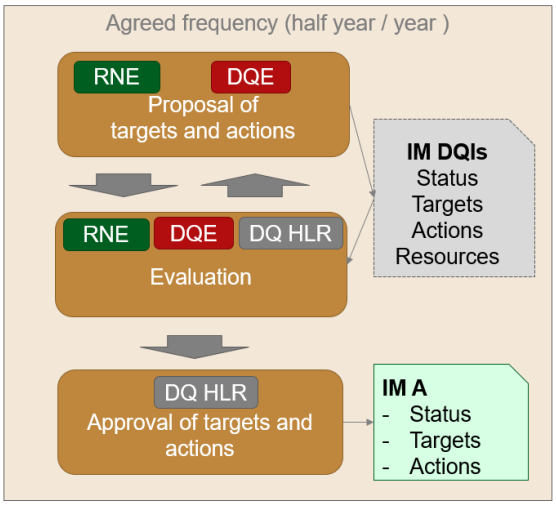


Figure 7: Steering of data quality targets

The monitoring, if the agreed targets were reached and agreed actions performed, is done on a half-year basis and results are regularly reported to RNE Managing level.

To support the above-mentioned process and to enable more detailed analyses, several different data quality reports are developed by RNE. These reports are available in the RNE Reporting tool (more details in chapter 0).

RNE Data Quality Expert Group

Main purpose of RNE Data Quality Expert Group (DQ EG) is to:

- Define, steer and monitor the complete Data Quality Management process,
- Identify global problems affecting the data quality and find appropriate measures to solve them
- Serve as platform for the exchange of experiences, e.g., in solving specific/local problems

The Data Quality Expert Group is led and coordinated by RNE Joint office and consists of Data Quality Experts from every single IM. The RFC TPM Coordinators, steering data quality on RCF level, are also welcomed members of this group.

Regular meetings will be held twice a year.

The goals of the meetings are:

» Discuss data quality status and reporting problems
» Identify the cause(s) of problems where RNE is not able to solve the problems with only one single IM.
» Discuss together problems that occur more often to be avoided in the future
» Update of Data Quality Handbook, its Annexes and procedures related to better management of data quality
» Evaluate and prepare the RNE Data Quality KPI – status, targets and actions to be taken

Based on the status and fulfilment of targets of each IM, the DQ Expert group on a yearly basis evaluates the overall fulfilment rate and overall level of Data Quality and set RNE targets and goals for the next year. This overall level of data quality is every year reported also to RNE Managing Board and General Assembly.

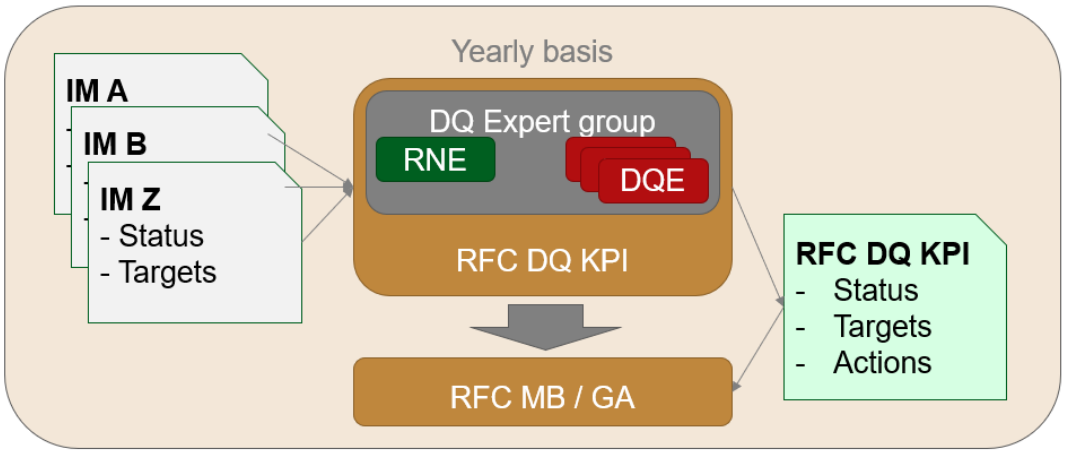


Figure 8: Overall level of data quality

RFC Data Quality Expert Level

Goal of the RFC DQ Expert level is to ensure, that its RFC Train Performance Management Working group (TPM WG):

- Is regularly informed about the data quality status along the RFCs and the commitment of RFC members to data quality processes
- Is regularly informed about problems and improvements related to data quality and influencing the content of RFC train performance management reports

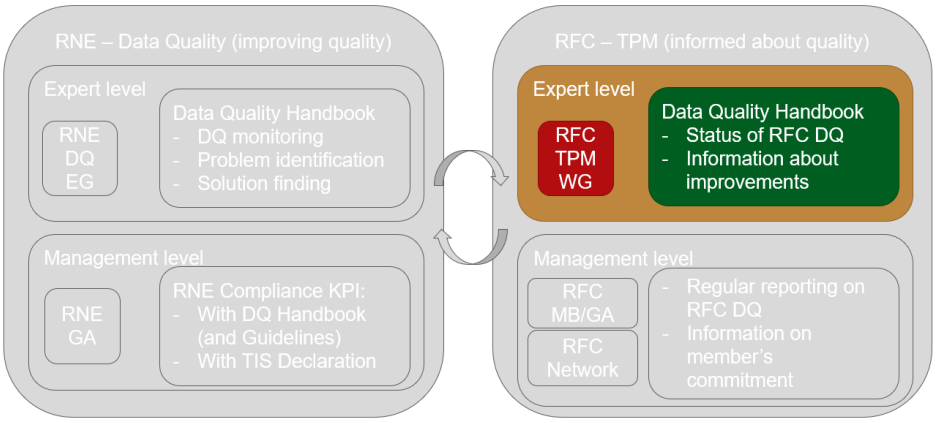


Figure 9: RFC Data Quality Expert level

The set of standard RFC data quality indicators to be monitored and reported to RFC TPM WGs and their calculation methodology is described in chapter 0. These indicators are calculated for each RFC separately and are included in the standard RFC Data Quality Report developed by RNE. The reports are available in the RNE Reporting tool. Each RFC shall execute and present this report at RFC TPM WG meeting at least on a half-year basis. The responsible person to execute and present the report within the RFC group shall be agreed upon by each RFC.

Each RFC TPM WG member is responsible to clarify all data quality-related questions with his national Data Quality Expert. Exceptionally, in case of an urgent need for detailed explanations related to data quality, national Data Quality Experts can be invited to participate on the RFC TPM WG meeting directly.

RFC data quality expert level is also regularly informed about the issues included in the Shortcoming list, having a direct and significant impact on the RFC data quality. Additionally, on yearly basis, RFC is informed about members' targets and actions.

RFC Data Quality Management level

The purpose of the RFC Data Quality Management level is to be regularly informed on the status of RFC data quality and involvement of the member IM's commitment to data quality management and improvement. RFC Data Quality Management level serves as a first escalation level in case of insufficient IM commitment.

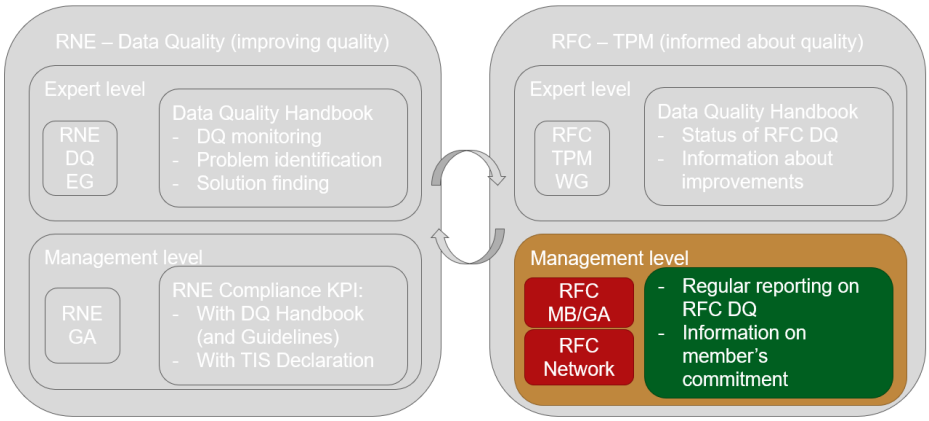


Figure 9: RFC Data Quality Management level

RFC Data Quality Management level is represented by:

- RFC Managing Board and/or General Assembly – for discussion on RFC specific data quality
- RFC Network – for discussion about the data quality impacting multiple RFCs.

RFC Managing Board and/or General Assembly

RFC Managing Board and or RFC General Assembly, serve as a first escalation level to increase the commitment of IMs involved in the RFC. Specific, RFC related, data quality reports shall be shown to RFC Managing level on a regular basis.

Reports show the status of RFC Data Quality Indicators and their development in time and can be found and executed directly in RNE reporting tool, together with a detailed description.

Additionally, RFC Managing Board and/or General Assembly shall be regularly informed about the engagement of their members in data quality management processes, especially via monitoring of members' target fulfilment.

RFC Network

At least on a yearly basis, detailed information about RFC data quality indicators and the commitment of RFC members to improve the data quality shall also be provided and reported to RFC Network.

RNE Data Quality Management Level

RNE Data Quality Management level is represented by RNE Managing Board and RNE General Assembly.

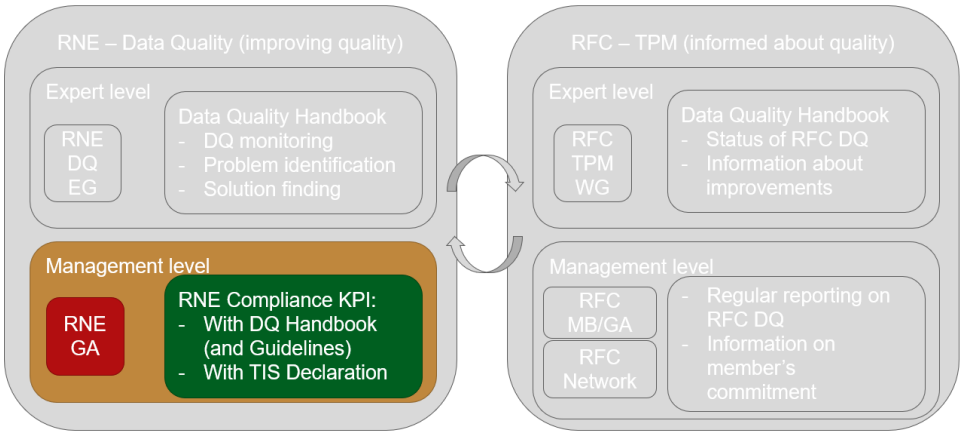


Figure 10: RNE Data Quality Management level

On a yearly basis, both RNE MB and GA shall be informed about Member's compliance with this Data Quality Handbook, via RNE Compliance KPIs and agree on corrective measures in case of insufficient compliance. Details about compliance monitoring can be found in Annex 5.

Overview of Tasks and Responsibilities

Each data quality level has its tasks and responsibilities, as described in the previous chapters.

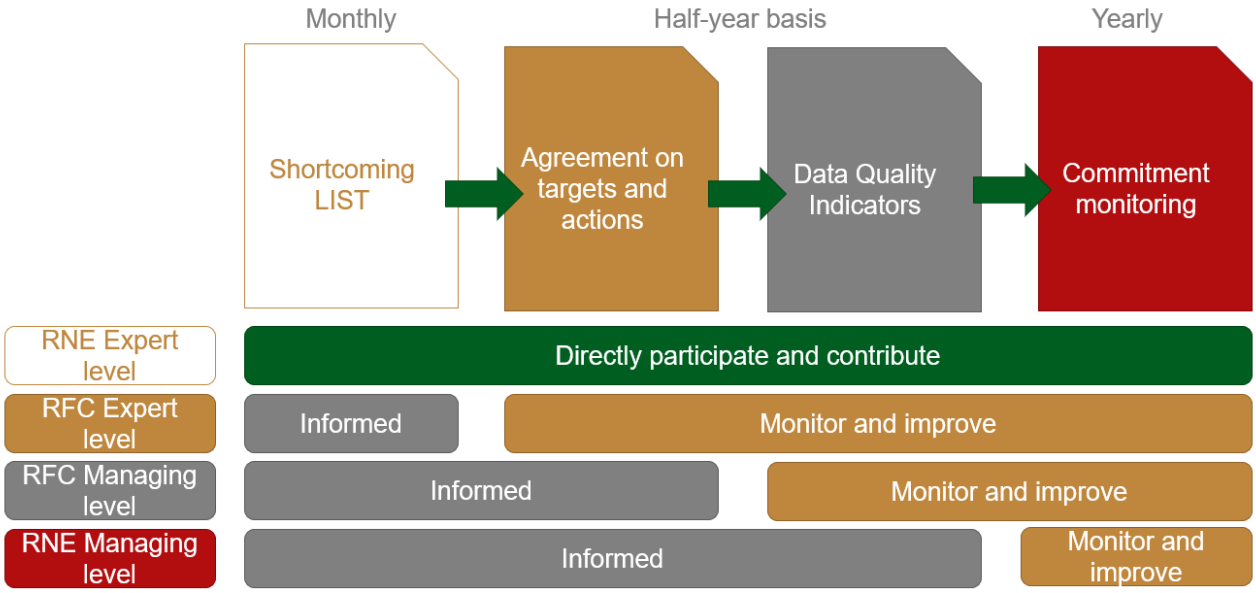


Figure 11: Overview of task responsibilities

Data Quality Measurement

Several different data quality indicators and data quality reports were agreed upon and developed and shall be used for the monitoring of data quality on all levels.

Following chapter provides the basic information about these indicators and reports.

Scope of Data Quality Indicators

The main goal of data quality indicators is to monitor the fulfilment of all requirements stated in the RNE Guidelines '*Basic requirements on TIS data delivery for reporting purposes'*.

Trains monitored

Based on the above-mentioned guidelines, the provision of information for all international trains (both freight and passenger) are mandatory. Therefore, on IM level, all below mentioned data quality indicators are measured separately for freight (national and international trains) and passenger trains.

For RFC level, the focus is only on trains related to the specific RFC. Always, the currently valid methodology for identification of RFC related trains is applied. The detailed description of this methodology can be provided on request.

Locations monitored

Based on the above-mentioned guidelines, the provision of information for all national locations is not required. Therefore, the main focus of data quality indicators is only on locations defined as mandatory in the above-mentioned guidelines, including also additional important points when relevant.

For RFC level, the focus is mostly on the RFC locations. Allocation of the location to a specific RFC is based on the RFC topology defined in the BIG DATA.

IM Data Quality Indicators

IM data quality indicators are focusing only on those data quality aspects, that are directly dependent on the single IM data delivery and thus their improvement is in the direct hands of the concerned IM.

Following IM Data Quality indicators are defined:

- Share of complete trains
  - representing the share of complete trains compared to all identified trains.
- Share of complete locations
  - representing the share of complete locations compared to all identified locations per train type.
- Share of documented delay minutes
  - representing the share of the delay minutes, for which the Train Delay Cause message was delivered to TIS compared to all delay minutes.

A detailed description of each indicator and precise calculation methodology is included in Annex 2.

RFC Data Quality Indicators

RFC data quality indicators are similar to IM Data Quality indicators but focus specifically on the data quality related to RFC trains in RFC locations. Additionally, they are also covering those data quality aspects, that are influenced by the cooperation and interface between several IMs and cannot be improved by single IM.

Following RFC Data Quality indicators were agreed upon:

- Share of complete RFC trains
  - representing the share of complete RFC trains compared to all identified RFC trains.
- Share of complete RFC locations
  - representing the share of complete RFC locations compared to all identified RFC locations
- Share of documented delay minutes of RFC trains in RFC locations
  - representing the share of the delay minutes of RFC trains, for which the Train Delay Cause message was delivered to TIS compared to all delay minutes.
- Share of linked trains in RFC borders
  - representing the share of trains linked in the border section compared to all trains

The detailed description of each indicator and precise calculation methodology is included in Annex 2.

Data Quality Reports

There are several standard data quality reports already developed and available in the RNE reporting tool (OAS), supporting the monitoring of data quality on each data quality level.

For each type of report, the standard content was agreed upon. Reports can have multiple parameters to be chosen from, therefore each user can additionally execute every report using different parameters to identify problems more easily and to help to propose possible solutions.

All Data Quality related reports are available in RNE Reporting tool together with a detailed report description explaining in detail the content of the report and calculation principles applied for all figures included in the report.

The detailed description of the current portfolio of Data Quality reports and reference to their descriptions can be found in Annex 3.

Data Quality tools

To support the monitoring of data quality, but especially the monitoring of improvements, the following tools shall be used :

- List of Shortcomings
- Data Quality Management tool

List of shortcomings

List of shortcomings is a web-based user interface to report and list all identified shortcomings, their impact on data quality, identification of needed potential corrective actions and proposal/evaluation of required resources.

List of shortcomings is the basic tool to be used for bilateral IM – RNE expert level on monthly basis.

The main functions of the List of Shortcomings:

- To be used as a ticketing system for identified issues – to keep a track of handling them
- Shortcomings can be listed by:
    - RNE experts
    - Data Quality Experts
    - Train Performance Management experts
    - TIS Change Control Board members
    - other

List of shortcomings can also be used as one of the sources for the identification of corrective actions to be included in Data Quality Management tool.

The detailed specification of List of shortcomings, including the basic description of main functions, can be found in Annex 4.

Data Quality Management tool

The Data Quality Management tool is a web-based user interface supporting the management and monitoring of Data Quality Targets and Actions.

Main purpose of the tool is to provide comprehensive overview of all agreed targets and actions and monitor the progress/development/improvements of data quality.

The tool is mostly relevant for bilateral IM-RNE usage on expert level and for interaction with IM Data Quality High-Level Representative, allocating resources.

Data Quality Management tool also serves as a Data Quality dashboard, providing:

- Actual status of Data Quality Indicators.
- Overview of agreed target values and timelines
- List of proposed actions and required resources
- Indication of agreement/commitment statement by Data Quality High-Level Representative

The detailed specification of Data Quality Management tool, including the basic description of the main functions and detailed instructions for its usage, can be found in Annex 5.

Implementation and Updates

The implementation of all requirements and processes stated in this Handbook on Management of data quality for train performance analysis shall be gradually started immediately after the RNE General Assembly approval on 31 May 2022.

All requirements and processes shall be fully applicable from January 2023 at the latest.

The requirements stated in this Handbook apply to all IMs already providing data to RNE TIS. In case a new IM is joining the TIS community and starts sending data to TIS, the full implementation of this Handbook might be exceptionally postponed based on mutual agreement between RNE and concerned IM, but not more than 1 calendar year.

The RNE Data Quality Expert group is responsible to keep this Handbook up-to-date and initiate the revision process when needed. Any change to the Handbook shall be presented and approved by RNE General Assembly.

Changes and updates of the Annexes to this Handbook are managed directly by the RNE Data Quality Expert group. Such changes are not subject for RNE General Assembly decision.

At the moment, the main focus of this handbook is on the IMs providing data to TIS. Other data providers (RUs, Terminals, new IMs etc.) are not yet covered by this version of the handbook but will be included in future updates, after first experiences with the implementation of the current handbook have been gained.

Annex 1: Commitment Monitoring

The overall commitment of all involved parties to this Handbook and RNE TIS Declaration is measured by an Overall Commitment Monitoring KPI per IM, as detailed in this annex.

This KPI is compounded by the sum of four terms. Each of these terms has equal weight (25%) in the calculation of the overall KPI and covers a different part of the process. These terms involved in the KPI calculation are indicated and explained below:

- **Term 1**: **Data Quality Expert Nominated**

According to this DQH, each IM shall nominate a Data Quality Expert to engage RNE Data Quality Expert level, participating in DQ management procedure, defining IM DQ targets and actions, and cooperating with the DQ Expert Group.

Therefore, if the IM has a DQ Expert nominated this term is considered complete and is valued at 25%. If the DQ Expert is not nominated the term is valued at 0%.

- **Term 2**: **Data Quality High-Level Representative Nominated**

This DQH requires each IM to appoint a Data Quality High-Level Representative. This DQ High-Level Representative is responsible for, on a half-year basis, approving targets and actions and committing to providing sufficient resources and efforts to achieve them.

Therefore, if the IM has a DQ High-Level Representative nominated, this term is considered complete and is valued at 25%. If the DQ High-Level Representative is not nominated, the term is valued at 0%.

It is established in the DQH that if a High-Level Representative is not nominated, the RNE GA representative shall take over the responsibility. However, for the overall commitment calculation, the nomination is considered complete if the RNE GA representative has been formally indicated as a DQ High-Level Representative.

- **Term 3**: **IM has a target for each KPI related to IM Data Quality Indicators**

Once KPIs are calculated for each IM, those shall have defined, for each KPI related to IM DQI, updated specific targets to be reached within a specified period. Each IM should have its targets according to its figures and perspectives.

The targets and time frame for achieving them must be set in the Data Quality Management Tool.

Therefore, if the IM has an updated target for each KPI related to DQI and a period to reach them defined in the Data Quality Management tool this term is considered complete and is valued at 25%. If not, the term is valued at 0%.

- **Term 4**: **IM has agreed to half-year actions to improve Data Quality**

To improve Data Quality, each IM should make an evaluation of its figures and define possible actions to enhance them. These actions must be established on a half-year basis, proposed by IM DQ Expert, with RNE Expert support, and validated by DQ High-Level Representative.

The approved actions and the period to achieve them shall be set in the Data Quality Management Tool.

Therefore, if the IM has updated actions to improve DQ defined in the Data Quality Management tool this term is considered complete and is valued at 25%. If not, the term is valued at 0%.

Term 4 is also considered complete if IM DQ Expert, RNE Expert, and DQ High-Level representative agree that there are no actions to be taken. This situation should only apply to IMs with a distinguished DQ performance.

Because some actions may take longer than six months or are not directly dependent on a single IM, this Term could be considered complete, as well, if there are no other six-month actions to be applied by the IM, and IM DQ Expert, RNE Expert and DQ High-level representative agree with the defined long-term actions.

Consequently, the Overall Commitment Monitoring KPI is in a range between 0 and 100% and is calculated as described below:

IM Overall Commitment Monitoring = Term 1 + Term 2 + Term 3 + Term 4

This commitment monitoring assessment method shall be evaluated by Data Quality Expert Group at least once a year and might be adapted according to experience and evolution of Data Quality process.

Annex 2: Data Quality Indicators

To monitor the fulfilment of all requirements stated in the RNE Guidelines Basic requirements on TIS data delivery for reporting purposes, several different data quality indicators were defined, developed, and can be found described in this annex.

Trains monitored

On IM level, all bellow mentioned data quality indicators are measured separately for freight international and national trains and for passenger international trains.

On RFC level, the focus is only on trains related to the specific RFC – so-called RFC related trains. Always, the currently valid methodology for identifying RFC related trains is applied. A detailed description of this methodology can be provided on request.

Only trains with at least one timetable (TT) or train running information (RI) available in TIS could be monitored and considered for data quality assessment.

National sections of trains with completely missing running information are not considered for Data Quality Indicators calculations, but information about these trains is expressed in the reports.

If a train run section is cancelled and that information (Path Section Notification Message) is sent to TIS. In this case, the train is only considered on the train run's evaluated (not cancelled) section.

Locations monitored

The main focus of data quality indicators is only on locations defined as mandatory in the above-mentioned guidelines, including also additional important points when relevant – further referred to as ‘important locations’.

The following location types are deemed as important locations:

- BORDER_PAIR_POINT – points defined as Train Identification Pair points in Border Management tool to identify the trains crossing the border in which the specified IM is involved
- BORDER_SECTION_POINT – points defined as starting/ending points of Measuring section defined in Border Management tool for borders in which the specified IM is involved
- BORDER_MEASURING_POINT – points defined as dwell time Measuring points in Border Management tool for borders in which the specified IM is involved
- LINKING_REGION_POINT – points defined as linking region location in TIS for linking regions where specified IM is involved
- OCTOPUS_POINT – points (defined in BIG DATA) which are connected to 3 or more segments – trains running in these points can run in 3 or more directions
- RFC_REPORTING_POINT – points defined as measuring points for specific RFC reports (points included in RFC reporting point list)
- TF_MEASURING_POINT – points defined as measuring points for specific Traffic Flow reports (defined in Report Management tool - Traffic Flow)
- RFC_BASIC_POINT_LIST – points defined in RFC Basic point list, used to identify the trains belonging to the specific RFC. If the new methodology for RFC train identification will be implemented and this list will not be relevant anymore, it will be excluded from the completeness measurement

- MASTER_STATIONS – the specific measurement concept on the reporting level was defined, to ensure the correct measurement in wider, usually shunting yards, areas, consisting of multiple substations, where timetable information might be available for different sublocations than running information. If such master stations are defined, for specific IM or RFC, the completeness measurement is done on the level of master station only, excluding the sublocations included in master-station

Master Station concept will always be considered in the Data Quality Indicators calculations and so Master Stations' sub-locations are excluded from the calculations. At a master station, information is considered complete if at least one running information from one of its sub-locations is provided.

Infrastructure Managers could request to exclude some locations if, due to their specificities, it is proven that is not possible to provide data properly. Each Infrastructure Manager is responsible for identifying the locations that should be excluded and requesting the exclusion to RNE, for validation and further exclusion.

Although Data Quality Indicators just reflect data provided in important locations, figures about the status of data concerning all nationwide locations are calculated and available in the IM reports.

TIS specificities

TIS stores information about train run as follows:

- The information about timetable time and running information is stored for each location and status (arrival, arrival at destination, departure, departure from origin, run-through)—this set is considered as one piece of information;
- If, for one location, two different statuses (either in timetable or running or both) are delivered, these statuses are treated as two separate pieces of information;
- TIS recognizes location only if at least one piece of information is delivered, either with a timetable or running information, or both.

Consequently, these specificities had been considered in the definition of Data Quality Indicators and their KPIs calculations.

IM Data Quality Indicators

The main focus of IM DQIs is only on those data quality aspects that are directly dependent on the single IM data delivery and thus their improvement is in the direct hands of the concerned IM.

The following IM DQIs were agreed upon:

- Share of complete trains
- Share of complete locations
- Share of documented minutes

Share of complete trains

Share of complete trains indicator is focusing on the measurement of completeness of information send to TIS per specified train types in important locations. This Data Quality Indicator is composed of two different kinds of KPIs, share of complete train per train type and share of information provided per train type.

Thus, for this Data Quality Indicator are considered six KPIs as presented below:

- Share of complete International Passenger trains in Important Locations;
- Share of information provided for International Passenger trains in Important Locations;
- Share of complete International Freight trains in Important Locations;
- Share of information provided for International Freight trains in Important Locations;
- Share of complete National Freight trains in Important Locations;
- Share of information provided for National Freight trains in Important Locations;

The principles considered for the calculations of share of complete trains KPIs are:

» Train completeness is measured in terms of the completeness of location information rather than all location. So, for one location with both arrival and departure statuses, two location information are considered.

» No distinction is made between the importance of timetable or train running information and so are both considered for the calculations as a piece of information.

» Missing Data from the timetable is not recalculated backwards from train running information, except for Master Stations.

KPI: Share of complete trains per train type

- These KPIs indicate the share of complete trains compared to all identified trains, per train type in important locations.

- **Complete train:** On the share of complete train per train type, a train is acknowledged as complete according to the share of information available (timetable and train running Information), and it is defined as **complete if all the expected information had been provided.**

- Calculations:

1. Share of complete information by train: $= \frac{number\ of\ location\ information\ provided}{number\ of\ location\ information\ expected}\ (\%)$;

2. Analysis of complete trains: $= share\ of\ complete\ information = 100\%$

3. KPI calculation: $= \frac{sum\ of\ complete\ trains}{sum\ of\ all\ trains}\ (\%)$;

Thus, calculations for KPIs per train type are presented in the table below:

Share of complete International Passenger trains in Important Locations	$= \frac{sum\ of\ complete\ International\ Pax\ Trains\ in\ Important\ locations}{sum\ of\ all\ International\ Pax\ Trains\ in\ Important\ Locations}$ (%)
Share of complete International Freight trains in Important Locations	$= \frac{sum\ of\ complete\ International\ Freight\ Trains\ in\ Important\ locations}{sum\ of\ all\ International\ Freight\ Trains\ in\ Important\ Locations}$ (%)
Share of complete National Freight trains in Important Locations	$= \frac{sum\ of\ complete\ National\ Freight\ Trains\ in\ Important\ locations}{sum\ of\ all\ National\ Freight\ Trains\ in\ Important\ Locations}$ (%)

KPI: Share of information provided per train type in important locations

- These KPIs indicate the share of trains' location information data (timetable and train running information) provided to TIS compared to all expected trains' location information data, per train type.

- Calculations:

1. Share of complete information by train: $= \frac{number\ of\ location\ info\ provided}{number\ of\ location\ info\ expected}$

2. KPI calculation: $= \frac{sum\ of\ trains\ location\ info\ provided}{sum\ of\ trains\ location\ info\ expected}$ (%)

Thus, calculations for KPIs per train type and considered location are presented in the table below:

Share of information provided for International Passenger trains in Important Locations	$= \frac{sum\ of\ International\ PAX\ Trains\ location\ info\ provided\ in\ Important\ locations}{sum\ of\ International\ PAX\ Trains\ location\ info\ expected\ in\ Important\ locations}$ (%)
Share of information provided for International Freight trains in Important Locations	$= \frac{sum\ of\ International\ Freight\ Trains\ location\ info\ provided\ in\ Important\ locations}{sum\ of\ International\ Freight\ Trains\ location\ info\ expected\ in\ Important\ locations}$ (%)
Share of information provided for National Freight trains in Important Locations	$= \frac{sum\ of\ National\ Freight\ Trains\ location\ info\ provided\ in\ Important\ locations}{sum\ of\ National\ Freight\ Trains\ location\ info\ expected\ in\ Important\ locations}$ (%)

Share of complete locations

Share of complete locations indicator is focusing on the measurement of completeness of information sent to TIS in reporting locations per specified train types.

Then, for this Data Quality Indicator three KPIs are considered, as presented below:

- Share of complete Important Locations for International Passenger trains;

- Share of complete Important Locations for International Freight trains;
- Share of complete Important Locations for National Freight trains;

The principles considered for the calculations of Share of complete locations KPIs are:

» Only important locations, for which at least one piece of information (timetable or train running information) is available for at least one train, are considered.

» Location completeness is measured in terms of the completeness of all train information for the location. (For example, to consider train data complete at one location with both arrival and departure statuses, two timetables and two train running information should be available.)

» **(90%) Location Completeness:** On the share of complete important locations, a location is acknowledged as complete according to the share of trains with complete data over all trains considered in that location, and it is defined as **complete if at least 90% of the trains have all data available.**

» Missing Data from the timetable is not recalculated backwards from train running information, except for Master Stations.

KPI: Share of Complete Important Locations per train type

- These KPIs indicate the share of (90%) complete important locations compared to all important locations, per train type.

- Calculations:

1. Location Completeness per train type: $= \frac{number\ of\ trains\ with\ complete\ data}{number\ of\ trains}$ (%)

2. Analysis of (90%) complete locations: $= Location\ completensess \geq 90\%$

3. KPI calculation: $= \frac{sum\ of\ (90\%)\ complete\ locations}{sum\ of\ considered\ locations}$ (%)

Thus, calculations for KPIs per train type are presented in the table below:

Share of Complete Important Locations for International Passenger Trains	$= \frac{sum\ of\ (90\%)\ complete\ locations\ for\ International\ PAX\ Trains}{sum\ of\ locations\ for\ International\ PAX\ Trains}$ (%)
Share of Complete Important Locations for International Freight Trains	$= \frac{sum\ of\ (90\%)\ complete\ locations\ for\ International\ Freight\ Trains}{sum\ of\ locations\ for\ International\ Freight\ Trains}$ (%)
Share of Complete Important Locations for National Freight Trains	$= \frac{sum\ of\ (90\%)\ complete\ locations\ for\ National\ Freight\ Trains}{sum\ of\ locations\ for\ National\ Freight\ Trains}$ (%)

Share of documented delay minutes

Share of documented delay minutes indicator is focusing on the share of the delays that the train delay cause message was delivered to TIS compared to all additional delays, per train type.

Then, for this Data Quality Indicator three KPIs are considered, as presented below:

- Share of documented delay minutes for International Passenger trains;
- Share of documented delay minutes for International Freight trains;
- Share of documented delay minutes for National Freight trains;

The principles considered for the calculations of Share of documented delay minutes KPIs are:

» Only the additional delay that occurred during the train running on the IM infrastructure is considered.

» To take into account the different IMs' trains documentation methodologies there are two options to consider delays:

- Every minute of additional delay is counted;
  or
- Excluding additional delays below 5 minutes;

The exclusion of additional delays below 5 minutes could be activated by the IM and should be done according to its national rules for documentation.

Independently of the option taken by the IM, all delays are stored at the database level.

» The calculations are based on the basic calculations in TIS.

» Documented minutes cannot be higher than the additional delay. If this occurs, only the value of the additional delay will be considered as documented.

KPI: Share of documented delay minutes per train type

- These KPIs indicate the share of total documented minutes compared to the total additional delay minutes, per train type.

- Calculations:

1. KPI calculation: $= \frac{sum\ of\ documented\ minutes}{sum\ of\ additional\ delay\ minutes}\ (\%)$

Thus, calculations for KPIs per train type are presented in the table below:

Share of Documented Delay Minutes for International Passenger Trains	$= \frac{sum\ of\ International\ PAX\ Trains\ Documented\ Minutes}{sum\ of\ International\ PAX\ Trains\ Additional\ Delay\ Minutes}\ (\%)$

Share of Documented Delay Minutes for International Freight Trains	$= \frac{sum\ of\ International\ Freight\ Trains\ Documented\ Minutes}{sum\ of\ International\ Freight\ Trains\ Additional\ Delay\ Minutes}\ (\%)$
Share of Documented Delay Minutes for National Freight Trains	$= \frac{sum\ of\ National\ Freight\ Trains\ Documented\ Minutes}{sum\ of\ National\ Freight\ Trains\ Additional\ Delay\ Minutes}\ (\%)$

RFC Data Quality Indicators

RFC Data Quality Indicators are similar to IM Data Quality Indicators but focus specifically on the data quality related to RFC trains in RFC locations. Additionally, they are also covering those data quality aspects, that are influenced by the cooperation and interface between several IMs and cannot be improved by single IM.

Following RFC DQ indicators were agreed upon:

- Share of complete RFC trains
- Share of complete RFC locations
- Share of documented minutes of RFC trains in RFC locations
- Share of linked trains in RFC borders

For RFC Data Quality Indicators, besides the principles related to trains and locations defined at the beginning of this annex, the following assumptions are applied:

» Always, the currently valid methodology for identification of RFC related trains is applied.

» Allocation of the location to a specific RFC is based on the RFC topology defined in the GeoEditor application.

» Only Important locations belonging to RFC in Geo-editor are contemplated.

» Whole train run is considered, but for calculations only information in RFC locations is counted.

» A train, to be considered, should belong to RFC and have at least one timetable or one train running information in one RFC important location.

» The master station concept is applied if the master station is allocated to RFC. A master station is allocated to a RFC if at least one sub-location belongs to RFC in GeoEditor topology.

Share of complete RFC trains

Share of complete RFC trains indicator is focusing on the measurement of completeness of information send to TIS for RFC trains in RFC important locations. This Data Quality Indicator is composed of two different kinds of KPIs, share of complete RFC trains in RFC locations and share of information provided for RFC trains in RFC locations.

Thus, for this Data Quality Indicator is considered two KPIs, as presented below:

- Share of complete RFC train in RFC locations;

- Share of information provided for RFC trains in RFC locations;

The principles considered for the calculations of Share of complete RFC train in RFC locations:

» Train completeness is measured in terms of the completeness of location information rather than all location. So, for one location with both arrival and departure statuses, two location information are considered.

» No distinction is made between the importance of timetable or train running information and so are both considered for the calculations as a piece of information.

» Missing Data from the timetable is not recalculated backwards from train running information, except for Master Stations.

KPI: Share of complete RFC trains in RFC locations

- These KPIs indicate the share of complete RFC trains compared to all RFC identified trains in RFC locations.

- **RFC Complete train:** On the share of complete RFC trains in RFC locations, a train is acknowledged as complete according to the share of information available (timetable and train running Information), and it is defined as **complete if all the information expected had been provided.**

- Calculations:

1. Share of complete information by RFC train in RFC location:

$$= \frac{number\ of\ location\ information\ provided}{number\ of\ location\ information\ expected}\ (\%)\ ;$$

2. Analysis of complete trains: $= share\ of\ complete\ information = 100\%$

3. **Share of complete RFC trains in RFC locations**:

$$= \frac{sum\ of\ complete\ RFC\ trains\ in\ RFC\ locations}{sum\ of\ all\ RFC\ trains\ in\ RFC\ locations}\ (\%)$$

Share of complete RFC locations

Share of complete RFC locations indicator is focusing on the measurement of completeness of information sent to TIS in RFC locations for RFC trains.

Then, for this Data Quality Indicator one KPI is considered:

- Share of complete RFC Important Locations for RFC trains;

The principles considered for the calculations of Share of complete RFC locations for RFC trains are:

» Only RFC important locations, for which at least one piece of information (TT or RI) is available for at least one train, are considered.

» RFC Location completeness is measured in terms of the completeness of all RFC train information for the location. (For example, to consider train data complete at one location with both arrival and departure statuses, two timetables and two train running information should be available.)

» **(90%) RFC Location Completeness:** On the share of complete RFC locations, a location is acknowledged as complete according to the share of RFC trains with complete data over all RFC trains considered in that location, and it is defined as **complete if at least 90% of the RFC trains have all data available.**

» Missing Data from the timetable is not recalculated backwards from train running information, except for Master Stations.

KPI: Share of Complete RFC Important Locations for RFC trains

- These KPIs indicate the share of (90%) complete RFC locations compared to all RFC locations for RFC trains.

- Calculations:

1. RFC Location Completeness for RFC trains:

$$= \frac{number\ of\ RFC\ trains\ with\ complete\ data}{number\ of\ RFC\ trains}\ (\%)$$

2. Analysis of (90%) complete RFC locations: $= RFC\ Location\ completensess \geq 90\%$

3. **Share of Complete RFC Important Locations for RFC trains**:

$$= \frac{sum\ of\ (90\%)\ complete\ RFC\ locations}{sum\ of\ considered\ RFC\ locations}\ (\%)$$

Share of documented delay minutes of RFC trains in RFC locations

Share of documented delay minutes of RFC trains in RFC locations indicator is focusing on the share of the delays that the train delay cause message was delivered to TIS compared to all additional delays, for RFC trains in RFC locations.

The principles considered for the calculations of the Share of documented delay minutes of RFC trains in RFC locations are:

» Only the additional delay that occurred during the RFC train running in the RFC locations is considered.

» To take into account the different IMs' trains documentation methodologies there are two options to consider delays:

- Every minute of additional delay is counted;
  or
- Excluding additional delays below 5 minutes;

The exclusion of additional delays below 5 minutes could be activated by the IM and should be done according to its national rules for documentation. The option of each IM is taken into account for RFC figures.

Independently of the option taken by each IM, all delays are stored at the database level.

» The calculations are based on the basic calculations in TIS.

» Documented minutes cannot be higher than the additional delay. If this occurs, only the value of the additional delay will be considered as documented.

KPI: Share of documented delay minutes of RFC trains in RFC locations

- This KPI indicates the share of total documented minutes for RFC trains in RFC locations compared to the total additional delay minutes for RFC trains in RFC locations.

- Calculations:

Share of documented delay minutes of RFC trains in RFC locations:

$$= \frac{sum\ of\ RFC\ trains\ documented\ minutes\ in\ RFC\ locations}{sum\ of\ RFC\ trains\ additional\ delay\ minutes\ in\ RFC\ locations}\ (\%)$$

Share of linked trains in RFC borders

In order to link two national train running sections into one international train, specific train linking procedures are implemented on TIS level. Detailed information about linking procedures is provided in RNE Guidelines Basic requirements on TIS data delivery for reporting purposes.

To measure the quality of the above-mentioned train linking procedure, a specific data quality indicator was agreed upon and is regularly monitored. As the success of the train linking procedure depends on the quality of data from both IMs involved in the border section, the DQI related to the share of linked trains in RFC borders is a border specific indicator (not IM specific). Therefore, this indicator is not part of the IM DQIs but is included in the RFC DQIs.

Measurement of the indicator is done for each specific border section, identified by the linking region defined in TIS.

KPI: Share of linked trains in RFC borders

- This KPI indicates the share of linked trains per RFC border.

- Calculations:

1. For each location defined in the linking region, the international trains are identified – for each side of the border separately. (if for one IM multiple linking regions are defined, the total number of international trains is calculated as the number of unique international trains in all linking region locations)

2. Number of linked trains is determined by the identification of international trains available on both sides of the border.

3. Exclusion of International trains with start/ending in linking region points and not included in linked trains.

- Trains excluded: Trains with the status “Arrival at Destination” or “Departure from Origin” either in the timetable or train running information in linking region point and not crossing the border.

4. Calculation of international trains that potentially should have been linked, for each side of the border. – Linked trains plus unlinked trains per IM at the border.

5. **Share of linked trains in RFC Border:**

$$= \frac{linked\ trains\ at\ the\ border}{(\mathbf{Higher\ value\ from})\ linked\ trains\ plus\ unlinked\ trains\ per\ IM\ at\ the\ border}\ (\%)$$

Thus, for each RFC border a KPI with just one figure is calculated and exposes the lowest share of linked trains comparing both IMs at the RFC border.

The KPIs considered and their respective calculation methods described in this annex shall be evaluated by Data Quality Expert Group at least once a year and might be adapted according to the IM presented figures and needs revealed by experience in the Data Quality process.

Annex 3: Data Quality Reports

This annex outlines Data Quality Reports developed in RNE Reporting Tool, Oracle Analytics Server (OAS), including a list of the current portfolio of Data Quality reports and references to their descriptions.

Data Quality Reports can be found in catalogue of OAS at web address:
https://reports.rne.eu/

Business Needs

Data Quality Reports are supporting management and monitoring of data quality related issue for each Infrastructure Manager (IM) respectively.

There are several standard data quality reports already developed and available in RNE reporting tool (OAS), supporting the monitoring of data quality on each data quality level.

Reports have multiple parameters to be chosen from, therefore each user can additionally execute every report using different parameters to identify problems more easily and to help to propose possible solutions.

All data quality related reports are available together with detailed report description explaining in detail the content of the report and calculation principles applied for all figures included in the report.

List of Reports

Below is a list of Data Quality available reports:

Data Quality IM Completeness
Calculated figures related to quality indicators including completeness of trains, locations, and undocumented minutes related to certain IM

Data Quality IM Completeness – raw data
Raw data of calculated figures in Data Quality IM Completeness Report

Data Quality IM Indicators
List of calculated KPIs as agreed by working group related to certain IM

Data Quality IM Indicators Overview
Overview of all KPIs for all IMs

Data Quality RFC Completeness
Calculated figures related to quality indicators including completeness of trains, locations, and undocumented minutes for IMs related to certain corridor

Data Quality RFC Indicators
List of calculated KPIs as agreed by working group related to certain corridor

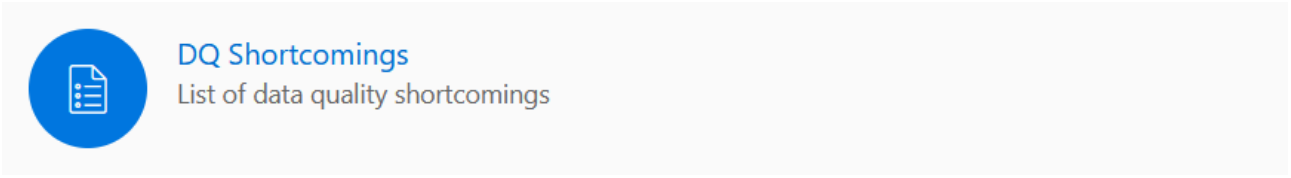
The list of reports shall be continuously updated according to the decisions taken by the Data Quality Expert Group related to Data Quality Indicators and respective KPIs and reports.

Annex 4: List of Shortcomings

Details related to specification of Data Quality Shortcomings developed within RNE Report Management Tool are described in this annex.

Specifically, this annex captures business, functional and technical specification that is required for mentioned tool.

This tool can be found as a part of RNE Reporting Management Tool at web address:
https://report-management.rne.eu/



Business needs

Data Quality Shortcomings is a web-based user interface to report and list all identified shortcomings, their impact on data quality, identification of needed potential corrective actions and proposal /evaluation of required resources.

This tool is mostly relevant for bilateral usage between IM and RailNetEurope (RNE) on expert level on monthly basis.

Data Quality Shortcomings is used to keep track of issues reported and as one of the sources for the identification of corrective actions to be included in Data Quality Management tool.

Functional and Technical Specification

Users and user rights

All Data Quality users can add new record to the list. This includes but not limited to:

- RNE experts
- Data Quality Experts
- Train Performance Management Experts
- TIS Change Control Board Members

All information is visible to all users. User can edit only created shortcoming and assign editing right to additional users.

Main page

Main page contains a list of all shortcomings.

Each shortcoming contains following elements:

- Title
- Description
- Priority defined as Low, Normal or High
- Responsible – multiple entries for companies and existing users

- Concerned parties – multiple entries for companies
- Status defined as open, progress, closed or on hold
- Expected fix date
- Ticket reference
- Comment
- User created
- User updated
- List of users who have right to edit a shortcoming
- Timestamp of creation
- Timestamp of last update

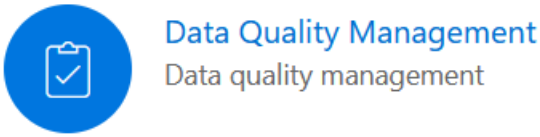
Shortcomings cannot be deleted from the tool.

Annex 5: Data Quality Management tool

This annex outlines details related to specification of Data Quality Management Tool developed within RNE Report Management Tool.

Specifically, this annex captures business, functional and technical specification, and detailed instructions that are required for mentioned tool.

This tool can be found as a part of RNE Reporting Management Tool at web address:
https://report-management.rne.eu/



Business Needs

Data Quality Management Tool is a web-based user interface supporting management and monitoring of data quality Targets and Actions for each Infrastructure Manager (IM) respectively.

Main purpose of the tool is to provide comprehensive overview of all agreed Targets and Actions and monitor progress, development, and improvements of data quality in general.

This tool is mostly relevant for bilateral usage between IM and RailNetEurope (RNE) on expert level and for interaction with IM Data Quality High-Level Representative for allocating necessary resources.

Data Quality Management tool also serves as a data quality dashboard, providing:

- Actual status of Data Quality Indicators (DQI)
- Overview of agreed target values and timelines
- List of proposed actions and required resources
- Indication of agreement/commitment by Data Quality High-Level Representative

Functional and Technical Specification

Users and user rights

Tool foresees three types of users:

1. Data Quality Standard (DQS)
2. Data Quality Expert (DQE)
3. Data Quality High Level (DQHL)

Each of the listed user type have specific rights:

Data Quality Standard User – can see everything on the main page of the tool and everything on individual page for each IM respectively.

Data Quality Expert User – can see everything on the main page of the tool and edit predefined sections of individual page for IM to which user is related.

Data Quality High Level User – can see everything on the main page of the tool and edit predefined sections of individual page for IM to which user is related.

Main page

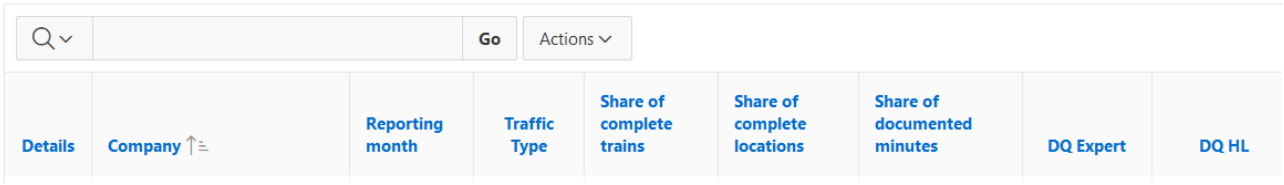
Main page contains a list of all IMs that provide data to RNE Train Information System (TIS) including Data Quality Indicators defined within RNE Data Quality Expert Group (DQEG) for each individual IM respectively. Value of DQI is related to reporting month.

In addition, main page lists DQE and DQHL users defined for individual IM respectively.

All information on this page is visible to all users.

DQE and DQHL users have access to “Edit” link for individual IM page to which user is related.

Each filed (column of a table) can be filtered to find specific information.



Individual IM page

Individual IM page displays DQI and is used to define and approve Targets and Actions that are planned to be achieved until specific validity period.

This page is visible to all DQ users related to specific IM.

Targets

For each target created, a unique ID is assigned called Target ID.
Target is defined by:

- DQI type
- Traffic type
- Target Value
- Validity
- Status

When status of a target is set to “Approved” by DQHL user, then change of validity is not allowed. If needed, new target can be created to be approved by DQHL.

User can define new target by adding new row in a table. Target deletion is possible only for rows with status “Pending”. For targets that are already saved only status can be changed to “Cancelled” if needed.

For one DQI multiple targets can be added depending on business case for short term goals, long term goals, etc.

Target status

Target can have several different statuses:

- Pending – by default when target is defined (version 0.0)
- To be approved – given manually by user (lead to change of version – 1.0)
- Approved – given by system if target is confirmed by DQHL user (version 2.0)
- Achieved – given by system if current DQ status value equals to target value within validity period (version 3.0)
- Not achieved – given by system if current value in validity month is below target value (version 4.0)
- Cancelled – given manually by user. Only “Pending” or “To be approved” status can be changed to “Cancelled” (version 5.0)

Any change of a target status is recorded including timestamp.

Target versioning

When new target is defined, target version 0.0 is given by the system. Every change made and saved by user leads to creation of new version.

Current DQ status and previous will change automatically within time and will not lead to new version creation.

Possible versioning scenarios:

- Status Pending: every change of validity, expert comments, and target status results in new version 0.0 – 0.99
- Status To be approved: every change of validity, expert comments results in new version 1.0 – 1.99
- Status Approved: every change of expert comments results in new version 2.0 – 2.99
    - As soon as status Approved is given, the changes in validity are not allowed anymore
- If status Achieved or Not achieved is given, no more changes are allowed – version 3.0 / 4.0
- If status Cancelled is given, version 5.0 is created, and no more changes are possible

All active targets are displayed ordered by validity period. If for one target multiple versions are available, then the target with highest version will be displayed.

Actions

List of actions is available to all users but editable only to DQE users.

DQE user can add as many rows as needed and edit all information. Each action has a unique ID. Every change to an action is recorded.

When new action is created it can be related to existing shortcoming from Data Quality Shortcomings Tool. Each action can have reference to Target ID defined in Targets section.

There is autocomplete offering existing shortcomings and targets.

In addition, there are more properties that can be defined for an action:

- Short name
- Detailed description
- Deadline until which the action should be performed
- Responsible entity which company/department shall take the action (i.e., RNE Joint Office, IM – TT department, IM – IT department, etc.)

- Required resources needed to fulfil the task; multiple can be specified (i.e., IT expert, traffic management expert, etc.)
- Estimated workload for each type of required resources i.e., hours needed, number of meetings, etc.
- Confirmation status with offered drop down menu including:
    - Pending (as default)
    - To be approved (given by DQE)
    - Approved (given by system)
    - Not approved (given by system)
    - Cancelled (allowed only for “Pending” and “To be approved”)
- Confirmation date of approval / not approval
- Completion status with drop down menu that can be given by user:
    - Pending (default)
    - In progress
    - Done
    - Completion date (date when status “Done” is given)

For actions with status “Approved”, only changes of Expert comment and Completion status are allowed

Rows are colour coded based on flag automatically (red, orange, green) for actions with “Approved” status:

- Red if completion status is “Pending” or “In progress” and deadline is in the past.
- Orange if completion status is “Pending” and deadline is closer than two months or completion status is “In progress” and deadline is same as current month.
- Green if completion status is “Done” or completion status is “Pending” or “In progress” and deadline in the future

Actions and targets for approval

Page displays list of actions and targets to DQHL user for approval.

List of actions and targets can be edited only by DQHL users.

List of Targets

View of targets is a same as for DQE user, but targets with status “To be approved”. DQHL user can edit approval filed as Approved or Not approved and/or write a comment. In addition, DQHL user can change the validity period.

Targets are colour coded as red, orange, green based on status same as in target section.

List of Actions

View of actions is a same as for DQE user, but targets with status “To be approved”. DQHL user can edit approval filed as Approved or Not approved and/or write a comment. In addition, DQHL user can change the validity period.

Targets are colour coded as red, orange, green based on status same as in actions section.