TECHNICAL REPORT

ISO/TR 21707

Intelligent transport systems — Integrated transport information, management and control — Data quality in ITS systems

Systèmes intelligents de transport (SIT) — Information des transports intégrée, gestion et commande — Qualité de données dans les systèmes SIT

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TR 21707 was prepared by Technical Committee ISO/TC 204, *Intelligent transport systems*.

Introduction

The publication and assessment of the quality of data that may be used by or exchanged between ITS systems and centres via integrated networks is vitally important. Without a knowledge of the quality of the data being exchanged, the usefulness of that data[1] is severely restricted, and whether it is fit for the intended purpose can not be established. In the worst case, it could lead to incorrect decisions being made due to wrong interpretations of the real occurrences upon which the data is based.

All data that does not have a stated quality should therefore be classed as unqualified and should be treated with appropriate caution.

Knowledge of the quality of data is relevant to all stages in the communication chain and is especially important where open systems are in place which have no knowledge of the recipient or ultimate use to which the data may be put. In particular, data quality is now a key issue for service providers who need to deliver accurate information to their clients. A high level of quality is needed for the information services to retain credibility with their customers (rebuilding trust is a very hard task).

Simply stating a measurement of quality associated with a piece of data does not in itself guarantee that the data source meets that quality. However, that is more a question of the monitoring and enforcement of service level agreements between data suppliers and data consumers and is outside the scope of this Technical Report.

This Technical Report sets out only a framework for the publication and assessment of data quality. The intention is that each type of data-application domain should have its own annex setting out the quality meta-data that are appropriate for their type of data and application.

---

1) Note that the term "data" is used throughout this document to mean the collective for data (plural).

Intelligent transport systems — Integrated transport information, management and control — Data quality in ITS systems

1 Scope

This Technical Report specifies a set of standard terminology for defining the quality of data being exchanged between data suppliers and data consumers in the ITS domain. This applies to Traffic and Travel Information Services and Traffic Management and Control Systems, specifically where open interfaces exist between systems. It may of course be applicable for other types of interfaces, including internal interfaces, but this Technical Report is aimed solely at open interfaces between systems.

This Technical Report identifies a set of parameters or meta-data such as accuracy, precision and timeliness, which can give a measure of the quality of the data exchanged and the overall service on an interface. Data quality is applicable to interfaces between any data supplier and data consumer, but is vitally important on open interfaces. It includes the quality of the service as a whole or any component part of the service that a supplying or publishing system can provide. For instance, this may give a measure of the availability and reliability of the data service in terms of uptime against downtime and the responsiveness of the service, or it may give a measure of the precision and accuracy of individual attributes in the published data.

In the majority of ITS applications, data is routinely exchanged between disparate systems. Where this data is being exchanged on a closed circuit between known senders and recipients, the parties concerned need to understand the quality of the data being exchanged and any resultant restrictions on its subsequent use by the recipient. In most cases, this is dealt with on a case-by-case basis and all parties to the agreement to exchange data will understand the quality parameters and restrictions.

However, transport and travel information is frequently being provided now via interfaces onto open networks for use by external users and it may not always be known from where this data has originated or for what purposes it is suitable. In these circumstances, a stated quality of the data becomes important and it is critical for users to understand the quality parameters so that accurate information can be derived from the data by itself or in combination with data from other sources.

Data quality meta-data includes the usual range of parameters normally associated with the measurement of quality such as accuracy, precision and timeliness of the data. However, there are other important quality meta-data such as ownership of the data. Ownership is important in many applications, and data suppliers may wish to restrict the usage of their data to certain classes of users. Measures of data quality may also be important in determining the relative monetary value of data in a commercial situation and so it is important that there is a common understanding of these measures.

It should be noted that, in the context of this Technical Report, data may be taken to be either raw data as initially collected, or as processed data, both of which may be made available via an interface to data consumers. The data consumer may be internal or external to the organization which is making the data available. Additionally, the data may be derived from real time data (e.g. live traffic event data, traffic measurement data or live camera images) or may be static data which has been derived and validated off-line (e.g. a location table defining a network). Measurements of data quality are of importance in all such cases.

This report is suitable for application to all open ITS interfaces in the Traffic and Travel Information Services domain and the Traffic Management and Control Systems domain.

2   Abbreviated terms

For the purposes of this document, the following abbreviated terms apply.

AE   Mean Absolute Error

AP   Availability Period

BC   Business Rules Coverage

CA   Calculation/Estimation Method

CM   Collection Method

CP   Calculation Period

DL   Standard Deviation Of Data Latency

ED   Error Standard Deviation

CV   Cross-Verified

DC   Data Correctness

DO   Data Owner

DP   Number of Decimal Places

DT   Data Type(s) Covered

DV   Data Validity Period

EM   Estimation/Simulation Model Identity

EP   Error Probability

ET   Equipment Type

FC   Physical Coverage

GC   Geographic Coverage

ITS   Intelligent Transport Systems

LR   Location Referencing Standard Identification

LT   Location Types

LV   Location Verification Standard

ME   Mean Error

ML   Mean Data Latency

MS   Measurement Source Identity

NP   Number of Data Points

OR   Data Owner's Original Reference

PC   Percentage Occurrence Coverage

RL   Reliability

RU   Restricted Use of Data

SF   Number of Significant Figures

SG   Service Grade

SL   Source of Location Data

SS   Spatial Data Set

TF   Mean Time Between Failures (MTBF)

TP   Time Precision

TR   Mean Time To Repair (MTTR)

TS   Data Time Stamping Regime

UI   Data Update Interval

UM   Data Update Mode

VP   Validation Process

3   General requirements

3.1   What is data quality?

Data quality is a slight misnomer since the "perception of quality" or "measurement of excellence" is not what we really mean here. These terms actually relate to the perception of quality by the data consumer and are terms used to assess the fitness for purpose of the received data. What we mean in this Technical Report by the term "data quality" is a set of meta-data which defines parameters relating to the supplied data or service that allows data consumers to make their own assessment as to whether the data is fit for their intended application. Different applications require different aspects of data quality and so it is not possible to say, for instance, that a data set with a reporting interval of one minute is of a higher quality than one with a reporting interval of 3 min. Only the data consumer can make this judgement of "perceived quality" since it must be based on the needs of their application (e.g. in terms of timeliness, accuracy, completeness, etc.).

3.2   What should a data quality standard define?

From the previous section it is clear that any standard for data quality should not be trying to define how measurements of excellence can be defined, but instead needs to identify what types of meta-data are appropriate and useful for a data supplier to provide and how this data may be structured and promulgated.

Different application and data domains within ITS may have very different requirements for data quality meta-data. It is therefore the intention that this data quality Technical Report specifies only a framework which each application and data domain can follow for identifying data quality requirements within their respective domain. Each ITS application and data domain will be required to define its own quality meta-data profile in a specific annex.

3.3   Data quality meta-data overview

Measurements of data quality are applicable to different levels within the structure of information flows across an interface.

At the lowest level, data quality meta-data is a measurement of the accuracy, precision or probability of correctness of any attribute within the data structure exchanged across an interface. For instance, this could be a measure of the accuracy of a location, a length of a queue or a timestamp, or it could be the probability of correctness of a severity estimate (selected from an enumerated list).

But data quality is also applicable at the higher level of data objects that flow across an interface. These data objects can be things like records defining an event or situation on a road, measurement of traffic flow or a camera image from a road video information system. The data quality meta-data which is applicable to these high level data objects is an assessment of the combined data quality of the individual attributes that go to make up the high level data object; for instance, does an accident event really exist or not.

Finally, an assessment of the quality of the data service as a whole or sub-parts of a data service that a supplier can offer to a data consumer is also an important measure. This is to do with the availability and reliability of the service as a whole and a definition of how well the data supplier covers the information in the live domain.

However, another way of classifying quality meta-data parameters is to determine whether they relate to the measurement of the quality of specific instances of data items, or whether they relate to the measurement of quality of data items, objects, the whole data service or parts of the whole service specified over a time period. The terms "instance data quality" and "generic data quality" are introduced for this purpose and can be expressed as follows.

⎯ *Instance data quality:*

Meta-data which gives a measure of quality for each specific instance of a data item. Each meta-data value is directly linked to an individual instance of data which flows across an ITS interface and either relates to an instance of a high level data object or to an individual attribute within a data object. This data would normally be promulgated along with the data itself and would therefore be included in the data model or schema of the published data. Each instance of a delivered data item will have its own value for these quality meta-data parameters.

⎯ *Generic data quality:*

Meta-data giving a measure of quality over time of a data service, parts of a data service, its component high level data objects or specific data items within those data objects. Different parts or components of a single data service would normally have different generic meta-data. It does not directly apply to individual instances of data since it is a measure over time. This meta-data can be provided off-line prior to any data consumer connecting to the service or sent separately from the data itself. It allows a pre-assessment of what can be expected from a service since it is a prediction of quality by the data supplier for a defined service period. Generic data quality meta-data are vitally important since they give a data consumer a clear idea of how useful the data might be in their intended application by defining predicted measurements of quality such as coverage, availability, veracity, timeliness, etc. They should allow a data consumer to assess one service against another. A data supplier providing different data services will need to define generic data quality meta-data for each service since it is likely that each will be different.

Of course these generic measurements of quality could be calculated retrospectively by an historical analysis of a data service. In fact, this may be how a data supplier derives some of the meta-data and it may be retrospectively derived in cases of dispute about service level agreements which relate to quality of data.

Clause 4 defines the different types of quality meta-data which should be considered for inclusion in a particular domain's data quality standard annex.

4 Data quality meta-data

Each type of quality meta-data is defined as a data quality object, each with a set of possible quality meta-data parameters. Data quality objects can be associated with a service as a whole, or with specific parts or entities within a data publication model and these would normally be delivered/published separately from the data feed. Also, data quality objects may be associated with specific groups or instances of data and may be delivered alongside the data to which it relates.

Clause 5 gives an indication as to which of these would normally be used as "instance" quality meta-data and which would normally be used as "generic" quality meta-data. Of course a system designer can put anything he likes in the data content model for the service and this Technical Report is not intended to discourage the inclusion of per-instance quality information where available.

Each application and data domain will need to define which of these quality data objects and meta-data parameters are applicable in their specific case.

Each quality meta-data parameter has been assigned a two letter code for optional use in interfacing to legacy systems where bandwidth may be limited.

4.1 Service completeness

The service completeness quality object contains meta-data parameters relating to a whole data service which should give a clear indication of how complete the coverage is of the stated service's domain. Depending on the type of service, there will be different sorts of completeness meta-data such as geographic based, data type based, event capture based, etc. For instance, a real time traffic event service should be able to clearly identify

— the geographical coverage of the service,

— the road types on which events are reported,

— the types of events which are reported, and

— a measure of the percentage of occurring significant events which would be expected to be reported.

The following meta-data parameters are applicable to the service completeness quality object.

— *Geographic coverage (GC):* A definition of the geographic area covered by the service. This may be specified in a variety of ways such as a spatial area by coordinate sets, by a regular geometric shape centred on a point (e.g. a circular area of given radius), as a named area (e.g. a fuzzy area such as a town's industrial area or a well-known tourist area), etc. When using coordinates to define geographic areas, it is important to define the coordinate system reference base and its time stamp.

— *Physical coverage (FC):* A definition of the physical coverage of the service. For instance, this might be a list of infrastructure items which are monitored by the service, such as lists of roads covered or types of sensors on specific roads that are covered.

— *Percentage occurrence coverage (PC):* A measure of the percentage of physical occurrences which the service is expected to report out of the total actual significant occurrences which occur in the defined coverage. A significant occurrence is something which meets the stated criteria for detection and reporting by the data source. A value of 100 % would imply that the service guarantees to report all significant occurrences in the defined coverage all the time.

— *Business rules coverage (BC):* A textual definition of the coverage in terms of specified business rules (e.g. a traffic flow service only covering vehicles of length > 5,5 m).

— *Data type(s) covered (DT):* A list of one or more types of data which the service will report. While this would typically be described by the data model/data format itself, there may also be cases where the same data model may be used for different data types, thereby justifying this attribute.

5

4.2   Service availability

A data supplier should be able to specify a set of meta-data which defines the availability of the service that a data consumer can expect (most probably derived from measurements of past performance), and demand if signed up to a service level agreement that references the service availability meta-data.

The following meta-data parameters are applicable to the service availability quality object:

— *Availability period (AP)*: The period during which the availability of the service to the specified standard is defined/guaranteed. This may be by time of day/week/month or other special periods (e.g. 24 h per day every day of the week, excluding national holiday periods).

— *Mean time to repair (TR)*: The mean time to repair (MTTR) the service fault and bring the service back into operation.

— *Mean time between failures (TF)*: The mean time between failures (MTBF) of the service.

The intrinsic availability of the service can be calculated as follows:

$$\text{Availability (intrinsic)} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$$

NOTE   Service availability meta-data may explicitly exclude communication problems external to the data suppliers organization (e.g. Internet links).

4.3   Service grade

A data supplier may wish to grade their data service to provide an overall assessment which may help data consumers decide the use to which the data may be put. If a particular ITS application domain wishes to use such a service grading system then some meaningful scales of grading must be defined by that domain. The meaning and justification of each grade will need to be defined by each application domain.

As an example, a Traffic Monitoring and Control application domain might define a set of service grades applicable to data services as shown in Table 1.

Table 1 — Service grades applicable to data services

Service Grade	Meaning
Safety critical	The data service is absolutely trustworthy and suitable for use in controlling equipments where lives would clearly be put at risk if the service provided erroneous data or failed. (e.g. control of traffic tidal flow systems or traffic access across a railway crossing, etc.)
High quality	The data service is very trustworthy and is suitable for controlling equipments which are critical to the operation of the road network.
Normal quality	The data service is generally reliable and is suitable for most control and monitoring operations, but should not be relied upon to control equipment affecting the critical operation of the road network or which control life-threatening situations.
Low quality	The data service is only reliable for monitoring applications and should not be used to control equipment.
Special quality	The data service can only be regarded as reliable within the bounds of special conditions defined for the service (e.g. weather conditions may affect the accuracy of certain measurement equipment).

Each application domain needs to decide whether such a grading system is useful, and if so, define a set of grades with explicit meanings and justifications which are specific to the application domain.

4.4 Veracity

The veracity quality object should contain meta-data which allows a data consumer to determine the expected correctness, truthfulness or error rate in the delivered data. This is applicable to high level data objects and attributes that are exchanged across an interface. From this metadata the data consumer should be able to make a judgement as to how reliable he should consider the delivered data to be. Different types of veracity meta-data may be applicable as follows.

— *Error probability (EP):* For a service providing dynamic data (i.e. reflecting real world occurrences) or static data (e.g. location table records) this is a measure or assessment of the probability of high level data objects (e.g. event or information records) which are delivered as part of the service being erroneous within the bounds of the accuracy and timeliness defined.

This is an important measure for a data consumer to be able to consider. For instance, if a data supplier is publishing event/situation information for a particular road (either in real time or historically), this would be their assessment as to whether events were likely to be erroneous or spurious. Obviously in this case it would be very dependent on the degree of certainty that the data supplier attaches to the source of their information (e.g. a service generating accident reports from only a single, non official source would probably be rated with a high probability of error). On the other hand, a traffic flow service which is derived directly from a loop which has good error checking and reporting would be rated with a low probability of error.

Error probability of a data object should be defined as 1 in 10^X, where "X" is explicitly specified. A default value for "X" of infinity could be valid for data items where they are guaranteed to be correct, such as for validated historical data.

For specific dynamic data attributes within a data object, other measures of correctness may be specified. Attributes may be expressing subjective or non-physical measurements (i.e. having a value domain defined by an enumerated list or free text) or be representing physical measurements. For these the following quality meta-data may be applicable [the first two are statistical measures relating normally to physical measurements and are illustrated in Figure 1 (which assumes a normal distribution of errors)].

— *Error standard deviation (ED)*: The expected standard deviation of the errors between reported values and the true values. This gives a measure of how far out a reported value is likely to be from the true value after taking into account any offset specified by the mean error (see following meta-data parameter).

— *Mean error (ME)*: The expected mean of the errors between reported values and the true values, giving a measure of whether the reported values are likely to be above or below the true values.

— *Mean absolute error (AE)*: The expected arithmetic mean of the non-signed errors.

— *Data correctness (DC)*: The probability of the data value being correct (within the bounds of any stated accuracy).

— *Reliability (RL)*: An indicator of "reliable" or "unreliable" subjectively determined by the data supplier (e.g. this could be derived from a sensor's fault monitoring system).

— *Cross-verified (CV)*: A verification flag indicating whether the data value has been cross verified from one or more additional sources.

It is up to the data consumer to decide whether to trust the reliability information provided by the data supplier.

For static type data which has been verified off-line (such as a static location referencing table), the following measure of veracity might be applicable.

Validation process (VP): A reference to a formal validation process which is applicable to that type of static data.

© ISO 2008 – All rights reserved
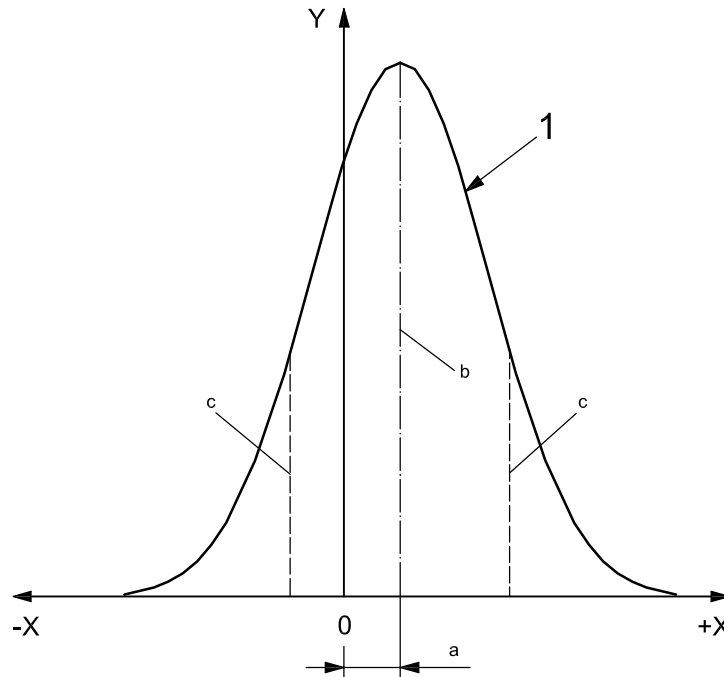
Key

1   distribution of errors between reported values and true values

Y   true values (zero errors)
–X   –ve error
+X   +ve error

a   Mean error.
b   Mean of errors.
c   Error standard deviation.

Figure 1 — Error deviation and mean

4.5   Precision

The precision quality object is applicable only to attributes exchanged across an interface which express a physical measurement (i.e. having a value domain which represents a physical measurement).

The applicable quality meta-data parameters are as follows.

⎯ *Number of significant figures (SF):* The number of significant figures to which the data is provided in the specified units.

⎯ *Number of decimal places (DP):* The number of decimal places to which the data is provided in the specified units.

⎯ *Time precision (TP):* In the case of a measurement of time, it should be indicated whether the data is provided to the nearest second, minute or hour.

4.6 Timeliness

The timeliness data quality object should contain a set of meta-data parameters relating to the timeliness of high level data objects that are delivered as part of a data service. Each type of high level data object within a data service may have its own specific set of timeliness meta-data which needs to be specified. Timeliness meta-data comprises the following.

— *Mean data latency (ML)*: This is the mean delay encountered between the capture of the incoming/live data and the output of the related occurrence/information to a data consumer. It is an important piece of quality information which provides a measure of the predicted staleness of high level data objects that are made available to the data consumer. Data latency should specify the latency period that will not be exceeded for instances of the high level data object which are exchanged as part of the data service.

— *Standard deviation of data latency (DL):* This is the standard deviation of the delay encountered between the capture of the incoming/live data and the output of the related occurrence/information to a data consumer.

— *Data update mode (UM)*: A data service needs to specify the regime for providing updates to the high level data objects which are published as part of the data service. The meta-data should state one or more of the following:

1) on occurrence (updates are made available to data consumers as and when they occur);

2) periodic (updates are made at regular intervals);

3) snapshot/on request (latest snapshot of data objects are provided when requested by a data consumer).

NOTE   The on occurrence and periodic regimes are dependent on a subscription type service being available.

— *Data update interval (UI)*: If periodic update mode is specified then the update interval should also be specified.

— *Data time stamping regime (TS)*: All instances of data objects should be time stamped with the time of capture or collection. This might be the time at which the raw data was collected by the sensor, the time the supplier system received the data or the time of computation if generated from many sources or over a period of time. Meta-data should be specified which defines what time stamping is applied to the data objects and how it relates to the collection or computation of the data object's content.

— *Data validity period (DV)*: Data objects may also be specified to have a validity period, after which the data objects can be considered to be no longer valid. Alternatively, data objects can be specified only to be snap shots of the real events/information, and as such can only be regarded as valid at the time of capture specified by the associated time stamp. For each type of high level data object, the expected data validity period is a useful piece of quality meta-data.

4.7 Location measurement

The location measurement quality object contains meta-data parameters relating to the referencing system used for locating the high level data objects. The applicable meta-data will be different depending on whether the location referencing uses on-the-fly referencing or static table referencing.

— *Location verification standard (LV)*: This is the standard against which the spatial data set used for referencing locations has been checked (e.g. reference to the Alert-C Forum validation tool). This is not applicable for on-the-fly location referencing systems.

— *Source of location information (SL)*: Identification of where the location information has been obtained from (e.g. road authority, police, civilian, national mapping agency, etc.).

The accuracy/precision of coordinate sets in a static location referencing table is a quality measure of the downloaded static table, while for on-the-fly location referencing accuracy should be given in the precision data quality object associated with the coordinate attributes.

For completeness, the following meta-data parameters have also been included here, although strictly speaking, these parameters actually define the location referencing format rather than specific data quality and may be included as part of the service description (e.g. as part of a directory of services entry). However, some may wish to classify them as data quality meta-data parameters.

— *Location referencing standard identification (LR)*: Identification of the location referencing standard used including any version and date (e.g. Alert-C, TPEG-Loc, AGORA-C).

— *Spatial data set (SS)*: Identification, version and date of the spatial data set used, for example the identity of an Alert-C national data table.

— *Location types (LT)*: This identifies the type of locations which will be used to locate data objects. The meta-data should state one or more of the following:

1) point;

2) linear (including itineraries/routes);

3) area (including named and geospatial definitions).

4.8   Measurement source

The measurement source quality object should identify how the data that is contained in the high level data objects is collected or measured. The meta-data may include the following.

— *Collection method (CM):* This can be one of four categories:

1) Manual: the data is provided by manual observation (e.g. manual survey, reporting or observation of CCTV etc.);

2) Raw: the data is provided as collected from a sensor, no post-processing has been conducted;

3) Calculated: the data provided is calculated from one or more raw data sources that may have been collected locally or by external systems (which can be named by the Information Publisher);

4) Estimated: the data provided is estimated from a mathematical model or simulation (which can be named by the information publisher.

— *Measurement source identity (MS)*: The identity of the source of the raw, calculated or estimated data may be provided.

— *Equipment type (ET)*: Identification of the type of equipment used to collect the raw data may be provided (e.g. loop, CCTV, automated weather station, etc.).

— *Estimation/simulation model identity (EM)*: Where the collection method is estimated via a mathematical model or simulation, the identity of the model may be provided.

— *Calculation/estimation method (CA)*: Where the collection method is calculated or estimated, the type of calculation or estimation method (e.g. arithmetic average over period or alpha-beta predictive filter) may be provided.

— *Number of data points (NP)*: Where the collection method is calculated or estimated, the number of data points used may be provided.

— *Calculation period (CP)*: Where the collection method is calculated or estimated, the period over which the values are determined may be provided.

4.9   Ownership

The ownership quality object should uniquely identify the owner of each set of high level data objects in a data service. This will often be the same as the publisher of the data, but not necessarily. The ownership of data is important to establish and maintain as data is passed from one system/organization to another. The terms and conditions which comprise the agreement between a data supplier and data consumer should specify whether the supplied data from the owner is for the sole use of the data consumer, or whether it can be passed on to third parties. It should be assumed that the owner retains copyright of the data unless explicitly stated otherwise in the agreement. The ownership meta-data therefore comprises the following.

⎯ *Data owner (DO):* Identity of data owner.

⎯ *Restricted use of data (RU)*: Whether use of data is restricted or unrestricted. If use is restricted, then the data consumer must refer to the terms and conditions in the agreement.

⎯ *Data owner's original reference (OR)* : The reference which the original owner assigned to the data object.

If no ownership meta-data is defined, then it should be assumed that copyright is owned by the data supplier and, unless stated otherwise in the terms and conditions agreed between supplier and consumer, that the data is for the sole use of the data consumer and should not be promulgated to third parties.

The *data owner's original reference* is important when data is forwarded via one or more intermediate nodes in a delivery chain and helps in data fusion processes and the traceability of the data. Where data sources have been fused, multiple occurrences of this meta-data may be used to trace the individual origins of the data.

5   Summary of data quality objects and their meta-data parameters

Table 2 lists the data quality objects, their possible meta-data parameters and associated optional short codes that should be considered by any application and data domain. It also indicates which of these would normally be expected to be used as "instance" or as "generic" quality meta-data. But system designers are not prohibited by this standard from including any relevant per-instance quality meta-data within or associated with their published data. It also may be appropriate for each ITS application domain to be more specific about the use of "instance" and "generic" quality meta-data.

"Instance" meta-data would normally be expected to accompany the delivered data, whilst "generic" meta-data would normally be delivered or published separately from the delivered data.

Table 2 — Data quality objects

Data quality object	Meta-data parameters (code)	Instance meta-data	Generic meta-data	How defined
Service completeness	Geographic Coverage (GC)		×	Text
	Physical Coverage (FC)		×	Text
	Percentage Occurrence Coverage (PC)		×	%
	Business Rules Coverage (BC)		×	Text
	Data Type(s) Covered (DT)		×	Text
Service availability	Availability Period (AP)		×	Text
	Mean Time To Repair (TR)		×	hh:mm
	Mean Time Between Failures (TF)		×	hh:mm
Service grade	Service Grade (SG)		×	Enumeration
Veracity	Error Probability (EP)		×	1 in 10^x
	Error Standard Deviation (ED)		×	Numeric
	Mean Error (ME)		×	Numeric
	Mean Absolute Error (AE)		×	Numeric
	Data Correctness (DC)	×	×	%
	Reliability (RL)	×		Boolean
	Cross-Verified (CV)	×		Boolean
	Validation Process (VP)		×	Text
Precision	Number of Significant Figures (SF)	×	×	Numeric
	Number of Decimal Places (DP)	×	×	Numeric
	Time Precision (TP)	×	×	Enumeration
Timeliness	Mean Data Latency (ML)		×	Numeric seconds
	Standard Deviation Of Data Latency (DL)		×	Numeric seconds
	Data Update Mode (UM)		×	Enumeration
	Data Update Interval (UI)		×	Numeric seconds
	Data Time Stamping Regime (TS)		×	Text
	Data Validity Period (DV)	×	×	hh:mm

Table 2 *(continued)*

Data quality object	Meta-data parameters (code)	Instance meta-data	Generic meta-data	How defined
Location measurement	Location Verification Standard (LV)		×	Text
	Source of Location Information (SL)		×	Text
	Location Referencing Standard Identification (LR)		×	Text
	Spatial Data Set (SS)	×	×	Text
	Location Types (LT)		×	Enumeration
Measurement source	Collection Method (CM)	×	×	Enumeration
	Measurement Source Identity (MS)	×	×	Text
	Equipment Type (ET)	×	×	Text
	Estimation/Simulation Model Identity (EM)		×	Text
	Calculation/Estimation Method (CA)		×	Text
	Number of Data Points (NP)		×	Numeric
	Calculation Period (CP)		×	hh:mm
Ownership	Data Owner (DO)	×	×	Text
	Restricted Use of Data (RU)	×	×	Boolean
	Data Owner's Original Reference (OR)	×	×	Text

ICS 03.220.20; 35.240.60

Price based on 13 pages