
**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*

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/TR 21707:2008](#)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>



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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/TR 21707:2008](https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2008

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword.....	iv
Introduction	v
1 Scope	1
2 Abbreviated terms	2
3 General requirements	3
3.1 What is data quality?	3
3.2 What should a data quality standard define?	3
3.3 Data quality meta-data overview	4
4 Data quality meta-data.....	5
4.1 Service completeness	5
4.2 Service availability.....	6
4.3 Service grade	6
4.4 Veracity	7
4.5 Precision	8
4.6 Timeliness	9
4.7 Location measurement.....	9
4.8 Measurement source	10
4.9 Ownership	11
5 Summary of data quality objects and their meta-data parameters	11

[ISO/TR 21707:2008](https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>

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*.

[ISO/TR 21707:2008](https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>

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 data¹⁾ 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.

iTeh STANDARD PREVIEW

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.

[ISO/TR 21707:2008](https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>

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

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/TR 21707:2008

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>

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

ITeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/TR 21707:2008](https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008)

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5cbacf70/iso-tr-21707-2008>

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

ITeh STANDARD PREVIEW

3 General requirements (standards.iteh.ai)

3.1 What is data quality?

ISO/TR 21707:2008

<https://standards.iteh.ai/catalog/standards/sist/cfe544de-896e-4748-a611-2a6c5ba6f70/iso-tr-21707-2008>

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.

STANDARD PREVIEW
(standards.iteh.ai)

— *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.