TECHNICAL REPORT

ISO/TR 25100

Second edition
2012-09-15

Intelligent transport systems — Systems architecture — Harmonization of ITS data concepts

Systèmes intelligents de transport — Architecture des systèmes — Harmonisation des concepts de données SIT

Reference number
ISO/TR 25100:2012(E)

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

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is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.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 25100 was prepared by Technical Committee ISO/TC 204, *Intelligent transport systems*.

This second edition cancels and replaces the first edition (ISO/TR 25100:2008). Clause 6 onwards has been technically revised.

Introduction

The objective of this Technical Report is to provide user guidance for the harmonisation of data concepts where there are similarities in definitions, including semantics.

Harmonisation has been discussed by several groups and some preliminary guidance and principles for the effective harmonisation of data concepts for Intelligent Transport Systems [ITS] has already emerged.

It should be clearly recognised that harmonisation is not essential for interoperability, which can usually be achieved given sufficient investment of knowledge and resources. Nevertheless this generally leads to duplication and other unnecessary, futile and even useless work being undertaken. This also assumes that there is an unlimited resource available to achieve the desired interoperability, whereas, in practice, time, budget and shortage of skilled resources often cause compromise. Additionally, interoperability in one aspect is sometimes achieved by the lack or loss of interoperability in another. Harmonisation is intended to reduce the nugatory work, increase efficiency and thereby reduce the incidence of errors and faults.

This Technical Report describes a proposed process for harmonisation of data concepts to arrive at preferred definitions for use in formal standards, specifications, technical reports and information models. The proposal is based on consideration of a harmonisation process used by international groups involved in transport and logistics information and control systems.

Harmonisation provides a means to improve efficiency and effectiveness of ITS, by helping to remove duplication, inefficiency, ambiguity and confusion, and thereby improve clarity, comprehension, safety and efficiency.

Intelligent transport systems — Systems architecture — Harmonization of ITS data concepts

1 Scope

This Technical Report provides guidance on the harmonisation of data concepts that are being managed by data registry and data dictionaries such as those described in ISO 14817:2002.

This Technical Report describes processes for harmonisation of such data concepts to arrive at preferred definitions for use in formal standards, specifications, technical reports and information models. It is based on consideration of a harmonisation process used by international groups involved in the ITS sector and in the wider sector of transport and logistics information and control systems.

2 Terms, definitions and abbreviated terms

2.1 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

2.1.1
attribute
data concept that represents a single property of an entity

2.1.2
data concept
definition of a kind of data representing a concept in the subject domain that can be identified with explicit boundaries and meaning and whose properties and behaviour all follow the same rules

NOTE This Technical Report assumes that data concepts, however they are represented, may have structure, such that individual property definitions are grouped into aggregate entities representing larger-grained concepts in the subject domain, and these entities may have relationships to one another; this basic idea is common to most description languages and metamodels including UML, XML and entity-relationship notations.

2.1.3
entity
data concept that may have attributes and relationships to other entities

NOTE This Technical Report follows common usage of the term “entity” where the words “entity kind” or “entity class” would be more accurate.

2.1.4
harmonisation of data concepts
process of reconciling differences in semantics, structure and syntax of similar data concepts

NOTE Harmonisation may include the establishment of a single pervasive definition for each data concept (i.e. standardization), but can also encompass flexible approaches in which definitions can be understood to grow closer without becoming identical.

2.1.5
ontology
rigorous conceptual schema representing the subject domain

2.1.6
relationship
property of a data concept that defines its relation to another data concept

2.1.7
standardization of data concepts
process of establishing a single standard definition for data concepts

2.1.8
taxonomy
classification scheme for a subject domain

2.2 Abbreviated terms

2.2.1
ACC
aggregate core component

2.2.2
ASCC
association core component

2.2.3
ASN.1
abstract syntax notation one

2.2.4
BCC
basic core component

2.2.5
BIE
business information entity

2.2.6
BRS
business requirement specification

2.2.7
CC
core component

2.2.8
CCC
change control committee [ISO 14817 (2002)]

2.2.9
CCTS
Core Components Technical Specification

2.2.10
CV
controlled vocabulary

2.2.11
DEN
Dictionary Entry Name

2.2.12
ETL
Extract, Transform and Load

2.2.13
IEC
International Electrotechnical Commission

2.2.14
ISO
International Organization for Standardization

2.2.15
ITS
intelligent transport systems

2.2.16
MOF
meta-object facility

2.2.17
QVT
queries views and transformations

2.2.18
RSM
requirements specification mapping

2.2.19
TBG17
'Trade and Business Processes Group working group 17', UN/CEFACT

2.2.20
TC
technical committee

2.2.21
TICS
transport information & control system

2.2.22
TIH
Travel Information Highway (UK)

2.2.23
UML
Unified Modeling Language (ISO/IEC 19501)

2.2.24
UN
United Nations

2.2.25
UN/CEFACT
United Nations Centre for Trade Facilitation and Electronic Business

2.2.26
WD
working draft

2.2.27
WG
working group

2.2.28
XML
eXtensible Markup Language

3 Background issues

Development of information systems and networks supporting business processes for transport and logistics frequently encounters multiple similar data concepts, any or all of which may be in widespread use. The need for harmonisation of these synonymous concepts has been acknowledged to enhance interoperability and reusability, but there are significant issues to be overcome.

Current approaches to achieve the data interoperability are principally to write ad-hoc data interface programs for each pair of communicating systems. Experience shows that development and maintenance of these programs is expensive in terms of both time and money. The total effort required increases with the square of the number of communicating systems.

3.1 Proprietary data concepts

The first issue is that many data concepts are proprietary or are deeply embedded in proprietary systems, which work well within their intended domain but are not freely accessible for broader use. There is an opportunity cost for a system whenever there is a similar but nevertheless separately defined and implemented concept in use in another domain that is not applied to the subject system.

3.2 Semantic differences

A second issue is where the concepts are subjects of widely used standards, but are not identical and have subtle semantic differences in their use. In this case the standards development organisations have generally been protective of their own approaches, based on concern about the cost of enforced changes on already deployed systems. This has resulted in diminished success in harmonisation processes (in the USA for example).

Semantic clashes are where similar or overlapping concepts are used with different detailed semantics in different standards.

3.3 Structural differences

Structural clashes are caused by the heterogeneity of representation which is possible with many techniques, such as XML representation. For example, using XML format the same concept can be expressed in several different ways. ISO 24531 provides assistance in these respects for the use of XML in the ITS sector.

XML Schema enables constraining of XML documents but this was designed for constraining the content of XML documents not for the conceptual representation. Within XML, structural clashes are mainly caused by the different usage of specific constructs, e.g. by a different usage of attributes rather than embedded elements or by expressing concepts in enumeration values.

Usually freely designed XML documents used for specific application purposes do not provide sufficient information about the semantics of the data. The semantics of XML elements used by web applications is hard-coded into the applications and is typically not available in machine processable form. This applies also to documents with available structural schemata (XML Schema), which in most cases define the syntactical structure of XML documents without explicit specification of their meaning.

Recording all standardised data using ASN.1, as specified in ISO 14813-6, provides assistance for defining structure and semantics, but of course does not prevent two independently designed structures from clashing.

Other forms of representation allow similar clashes to exist.

3.4 Difficulty of application of existing data concepts

When addressing a new application domain there should be a desire to reuse concepts that already exist as proprietary or open standards, but the mechanism to render them usable may be unclear. This generally results from semantic differences or uncertainty in the application of the concept, or because significant domain knowledge is required for the successful reuse of a data concept from a different domain. It can appear easier to an engineer to design a new concept rather than verify that an existing one is exactly suitable. Existing concepts tend to come within structures that are not optimal for further new applications, and unnecessary surplus structure discourages re-use.

3.5 Report of investigation

'Harmonisation' is often touted as the means to resolve these issues, but has been much more difficult to achieve than expected. This Technical Report is based on an on-going investigation being carried out on behalf of ISO/TC 204 working group WG1 (Intelligent Transport Systems [ITS] Architecture, Taxonomy, Terminology and data modelling), into various approaches used for harmonisation. This Technical Report presents tentative conclusions regarding the effectiveness of the approaches for general use in intelligent transport systems, and the wider sector of transport and logistics.

4 Harmonisation – General discussion

4.1 Introduction to harmonisation

Harmonisation in this context is the process of resolving differences in data definitions that have semantic overlaps. However, successful achievement of the harmonisation process remains a problem in many areas. Members of ISO/TC 204 WG1 have been considering this matter for some time and propose solutions to the requirement for effective harmonisation at syntactic, relationship and semantic levels. These solutions for harmonisation are provided in this Technical Report.

Progress in this respect has also been achieved in the 'United Nations Office for Trade Facilitation and Electronic Business' [UN/CEFACT] by the 'Trade and Business Processes Working Group' [TBG], specifically TBG17, as discussed in a subsequent section.

Harmonisation is easier to achieve if a single organisation owns all of the systems or specifications being harmonised. Harmonisation is particularly difficult in a mature domain where there are already established implementations and standards but no single controlling authority to enforce the use of one particular standard. Nevertheless even within a loosely aligned community, a harmonisation process can still be valuable in signalling preferred representations and providing aids to translation or migration.

4.2 Illustration of the need for harmonisation

It is helpful to consider the nature of the problem to be resolved. Take for example the need for integrated use of travel information in an advanced national traveller information service. One class of information for the traveller information system will be timetables for various travel services. To take an example from Australia where two timetables are to be merged but the times of service departure are expressed differently:

— Travel service A departure time format: local time in New South Wales (time zone universal coordinated time + 10 hours) 12-hour clock, subject to daylight savings time (Concept A);

— Travel service B departure time format: 24 hour clock based on Western Australia (time zone universal coordinated time + 8hrs) and not subject to daylight saving time (Concept B).

Of course, if the travel service were totally local, and travellers had no mobility, the only criteria would be local custom. However, as the object of travel is mobility, and we may expect one traveller to move from one locality to another, we may expect one travel provider to be providing travel information to traveller information systems elsewhere, and in these days of Internet we may also expect direct enquiries from elsewhere; there is therefore a significant benefit to be gained from harmonisation. An analyst can understand that both services have an implementation of the same underlying concept – the departure time of a travel service – and can therefore take steps to harmonise. It will be apparent that there is a need for a series of conversions and business rules to be applied to arrive at a compatible format, which could be in either of the proponent formats or a third (preferred) option could be the use of a standard time such as UTC with the conversion to the time format as preferred by the person making the inquiry (query) to be made at the time of a query.

4.3 Challenges in harmonisation

For a single underlying domain concept, there are many types of difference that can arise between the expressions of that concept in two different systems.

The following example is from a European project (Harmonise) for the 'Conceptual Normalisation of XML Data for Interoperability in Tourism'.

This project studies problems in using XML data in the tourist industry, and while much of its harmonisation resolution is very specific to XML it provides a methodology that in process (if not in detail) is similar to those described in this Technical Report, and provides some good examples of the problems involved. These are shown clearly in Tables 1 and 2 below.

Table 1 — Sample of differences

Different naming	PostCode vs. PostalCode
Different position	Postcode in Address rather than in ContactInfo
Different scope	TelephonePrefix and TelephoneNumber separated vs. as single concept

In the wider industry there is a need to resolve these differences to achieve interoperability. Harmonisation is possible because an observer can still see that, for example, “postcode” and “postalcode” are still expressions of the same domain concept.

The example in Table 2 shows three valid but different ways of expressing the concept *PostalCode* in XML.

Table 2 — Structural heterogeneity of XML

```
<ContactInformation>
    <Address PostalCode="X-1220">
    Wannaby Street 59, Dreamtown</Address>
</ContactInformation>
```

```
<ContactInformation>
    <Address>
        <Street>Wannaby Street 59</Street>
        <City>Dreamtown</City>
        <PostalCode>X-1220</PostalCode>
    </Address>
</ContactInformation>
```

```
<ContactInformation>
```



```
    <Address>
        Wannaby Street 59,
        <PostalCode>X-1220</PostalCode>
        Dreamtown
    </Address>
</ContactInformation>
```

The post code example shows some differences that can arise in the expression of one attribute within aggregate entities. In general there will be a *set* of entities *partially* corresponding to another *set* of entities.

Even where two systems or specifications apparently have similar scope when viewed at a high level, there may be entities present in one system that are entirely missing in the other.

In an example from highway location referencing in the UK, one data model included the following three concepts:

while another system used:

There was an approximate semantic equivalence between “RoadSection” and “Section”, and between “Section_LRP” (which stands for Location Reference Point) and “RoadsidePoint”, but there was no equivalent for “Link” – due to differing requirements and business rules about segmentation of the road network. Any harmonisation between the two models has to resolve the issue of how to resolve the relationship of the “Sections” to the “Points”, which in the second model is direct but in the first model is through the intermediate concept of “Link”. Harmonisation of the “Section” and “Link” entities would also have to resolve the differences in business rules.

Harmonisation has thus to deal with issues at a semantic level, at a structural level, and at a syntactic level. Every part of a data model could potentially vary from system to system even though the same concepts were being described. These parts will include names, attributes, relationships, the boundaries of structures and datatypes. And although the scope of harmonisation is for semantically related concepts, the detailed semantics and business rules may differ and therefore also require resolution.

4.4 Harmonisation processes

4.4.1 Harmonisation contrasted with standardization

NOTE This Technical Report uses the definitions in Clause 2 for these terms.

A well established approach to deal with the issues raised above is to standardize on a single format to be used in all applications. This can be very effective. However there may be forces which make complete standardization difficult. Often the same concepts occur in different business contexts, but the realisation of a concept that suits one business context may be very unsuitable for another business context. Harmonisation processes recognise this by trying to reconcile differences to a practical level without always enforcing the use of a single standard realisation of each concept in all business contexts.

The processes are clearly related, for example the outputs of harmonisation may subsequently be used as standards, but harmonisation is always focussed on reconciling differences across more than one set of definitions, whereas standardization may simply be the establishment of a single set of definitions.