TECHNICAL REPORT

ISO/TR 25100

First edition
2008-04-01

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 25100:2008(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is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.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*.

Introduction

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

Harmonization has been discussed by several groups and there has already emerged some preliminary guidance and principles for the effective harmonization of data concepts for intelligent transport systems [ITS].

It should be clearly recognized that harmonization 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, even futile work being undertaken. This also assumes that there are unlimited resources available to achieve the desired interoperability, whereas, in practice, time, budget and shortage of skilled personnel often cause compromise. Additionally, interoperability in one aspect is sometimes achieved by the lack or loss of interoperability in another. Harmonization is intended to reduce the inconsequential work, increase efficiency and thereby reduce the incidence of errors and faults.

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

Harmonization provides a means by which 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

The scope of this Technical Report is the harmonization of data concepts that are being managed by data registries and data dictionaries such as those described in ISO 14817:2002.

2 Terms and definitions

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

core component
aggregate information entities and the embedded entities within them

data concept
data dictionary structures defined in this Technical Report (i.e. object class, property, value domain, data element concept, data element, data frame, message, interface dialogue, association) referring to abstractions or things in the natural world that can be identified with explicit boundaries and meaning and whose properties and behaviour all follow the same rules

harmonization
process to resolve differences in synonymous terminology when expressed precisely in syntactic form

3 Abbreviated terms

ACC	aggregate core components
CEFACT	United Nations Centre for Trade Facilitation and Electronic Business
CCC	change control committee
CCTS	core components technical specification
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITS	intelligent transport systems
TBG17	UN/CEFACT Trade and Business Processes Group working group 17
TC	technical committee
TICS	transport information and control system
TIH	transport information highway (UK)
UML	unified modeling language
UN	United Nations

UTC	coordinated universal time
WD	working draft
WG	working group

4 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 harmonization 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. If you consider this problem and its cost implications further, it can be seen that the total effort required increases with the square of the number of communicating systems.

4.1 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.

4.2 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 organizations (SDO) have generally been protective of their own approaches out of concern about the cost of enforced changes to already deployed systems. This has resulted in diminished success in harmonization processes (in the USA for example).

Semantic clashes are clashes between concepts of different standards, or more precisely, between specific conceptual models or ontologies behind different standards. Typical semantic clashes are completely different concepts, different naming of concepts or different granularity.

4.3 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, *Intelligent transport systems — System architecture, taxonomy and terminology — Using XML in ITS standards, data registries and data dictionaries, 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 the conceptual representation. Within XML, structural clashes are mainly caused by the different usage of specific constructs, for example 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 the most cases define the syntactical structure of XML documents without unified implicit representation of their meaning.

Other forms of representation, with the possible exception of ASN.1, allow similar clashes to exist.

4.4 Difficulty of application of existing data concepts

A further issue is that there is the requirement in addressing a new application domain to reuse concepts that already exist as proprietary or open standards but for which the mechanism to render them usable is unclear. This generally results from semantical 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.

4.5 Report of investigation

Harmonization 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/WG 1 [*Intelligent transport systems, Architecture*] into various approaches used for harmonization. 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.

5 Harmonization — General discussion

5.1 Introduction to harmonization

Harmonization is a process to resolve differences in synonymous terminology when expressed precisely in syntactic form. However, successful achievement of the harmonization process remains a problem in many areas. Members of ISO/TC 204/WG 1 have been considering this matter for some time and propose solutions to the requirement for effective harmonization at syntactic, relationship and semantic levels. These solutions are provided in this Technical Report for harmonization.

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 below.

5.2 Illustration of the need for harmonization

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 (NTIS). 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 UTC+10 h or UTC+11 h), 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 UTC+8 h) 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, we may expect a traveller to move from one locality to another, or a travel provider to be providing travel information to traveller information systems elsewhere, or, in these days of Internet, we may expect direct enquiries from elsewhere. There is, therefore, a significant benefit to be gained from harmonization. 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. Alternatively, a third (preferred) option could be the use of a standard time such as UTC (Concept P) with the conversion to the time format as preferred by the person making the enquiry (query) to be made at the time of a query.

A second example can be taken from a European project (Harmonise) for the Conceptual Normalization of XML Data for Interoperability in Tourism. This project studies problems in using XML data in the tourist industry and, while much of its harmonization resolution is very specific to XML, it provides a methodology that in process (if not in detail) is similar to that proposed in this Technical Report, and provides some good examples of the problems involved. These are shown clearly in Table 1 and Table 2.

Table 1 — Sample of semantic clashes

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

The example in Table 2 shows three technically correct, according to the standards, 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>
```


Harmonization has thus to deal with issues at a semantic level, at a structural level, and at a content level.

5.3 Harmonization scenarios in data modelling terms

The essential process of harmonization is to resolve the differences between two or more data concepts in an agreed manner that has wider usage than merely the original data concepts. In simple terms this is shown in Figure 1:

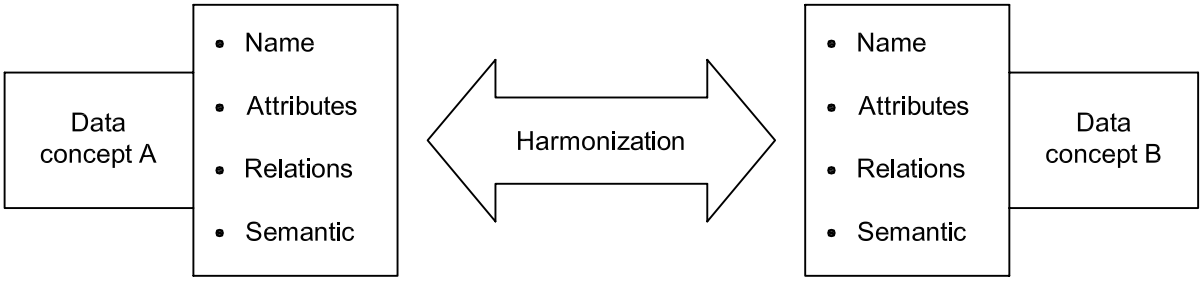

Figure 1 — Simple illustration of harmonization

Harmonization can then be defined as follows:

> For any pair of data concepts (*A,B*) harmonization is the selection of preferred concept *P* based on the attributes, relationships and semantics for individual data concepts *A* and *B*
>
> $P = h(A,B)$ where *h* is the harmonization preference function.

For the example above, the following scenarios apply to the harmonization function *h*.

5.3.1 Scenario 1: *h*{ [name(A)], [name(B)] } ⇒ name(P)

Harmonization shall generate a preferred name for the preferred data concept.

However, in generating the 'name' a process of 'conceptual normalization' (source: 'Harmonise' project) is an intrinsic part of this first part of the harmonization process.

By first agreeing a preferred name, an agreed basis for the object of the data concept is agreed at a highly abstracted level, without getting concerned at this early stage with the structure of the concept.

The separation between semantic and structural clashes indicates the need for a distinction between corresponding steps in the overall transformation process. This step enables a separation of semantic mapping (resolution of the semantic clashes) from the concrete physical representation of data being transformed. If different physical representations are used in the future, the semantic mapping definitions will still remain valid.

NOTE Whereas project 'Harmonise' only dealt with XML schema and proposed taking the conceptual normalization not only to include the name but also to 'provide a unified human and machine understandable description of the concepts of local systems and relations among them', both human and machine understandable description are only possible where there is a high degree of consensus concerning the form of the data concept (i.e. in the project 'Hamonise' context it is already a precondition that it is an XML schema) prior to commencement of the harmonization process. The process recommended in this deliverable takes a more pragmatic approach, and one that, the authors believe will overcome some of the potential weaknesses in agreeing schemata too early at a conceptual level.

5.3.2 Scenario 2: *h*{ [attribute$_n$(A)], [attribute$_n$(B)] } ⇒ attribute$_n$(P)

Each attribute of concept A shall be harmonized with each corresponding attribute of concept B to produce a corresponding attribute of preferred concept P.

Where the complexity of the semantical use of the attribute precludes direct harmonization, each attribute shall be expanded to a new data concept and the harmonization function applied iteratively until resolved (as discussed below).

5.3.3 Scenario 3: *h*[rel(A,X), rel(B,Y)] ⇒ rel(A,X), rel(B,X), rel(A,Y), rel(B,Y), rel(A,B)

Each relationship between the proponent concept and another concept shall be replicated for the other proponent concept. A relationship shall also be defined between the proponent concepts.