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)

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is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.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

Harmonisation increases interoperability, which need not be all or nothing: for example some parts of harmonised specifications could produce limited interoperability, perhaps with some special work needed in order to exchange data successfully; the better the alignment between the specifications the less special work is needed.

4.4.2 Harmonisation by recognising underlying concepts

Harmonisation was possible and worthwhile in the above examples because the data models or other data specifications were an expression of overlapping concepts from the subject domain. The scope and structure of each model has been influenced by specific business contexts and other contextual factors, but each does reflect concepts from the subject domain. Looking at two sets of data with overlapping semantics but different formats, an analyst is able to attempt to understand similarities because the descriptions show that the semantics overlap – both models are talking about the same domain concepts. The analyst is able to identify semantic similarities because he/she has a mental reference model of the subject domain, and can identify that, despite the differences that may exist, the data definitions in each set are a representation of concepts in the subject domain.

In simple cases it is possible to proceed to harmonisation directly without making the background reference model explicit. In these cases one data model will be changed to make it more similar to another. The extreme example of this is complete standardization, where the two contexts come to use exactly the same model.

In more complex situations there is value in making the background reference model explicit in the form of *ontology,* i.e. a rigorous conceptual schema representing the subject domain. The difference in nature between the background ontology and the data specifications undergoing harmonisation must be stressed. Admittedly, the data specifications may already be seen as attempts at rigorous expressions of the subject domain, but they are shaped by their specific business contexts. The ontology used in harmonisation should be more independent of business context. A harmonisation process can then use this reference ontology in the understanding and also the recording of the similarities and differences between different data specifications. These ideas are developed more fully in the UN/CEFACT Core Components approach, and the refined harmonisation approach used by the UK Highways Agency; both are described in Clause 5.

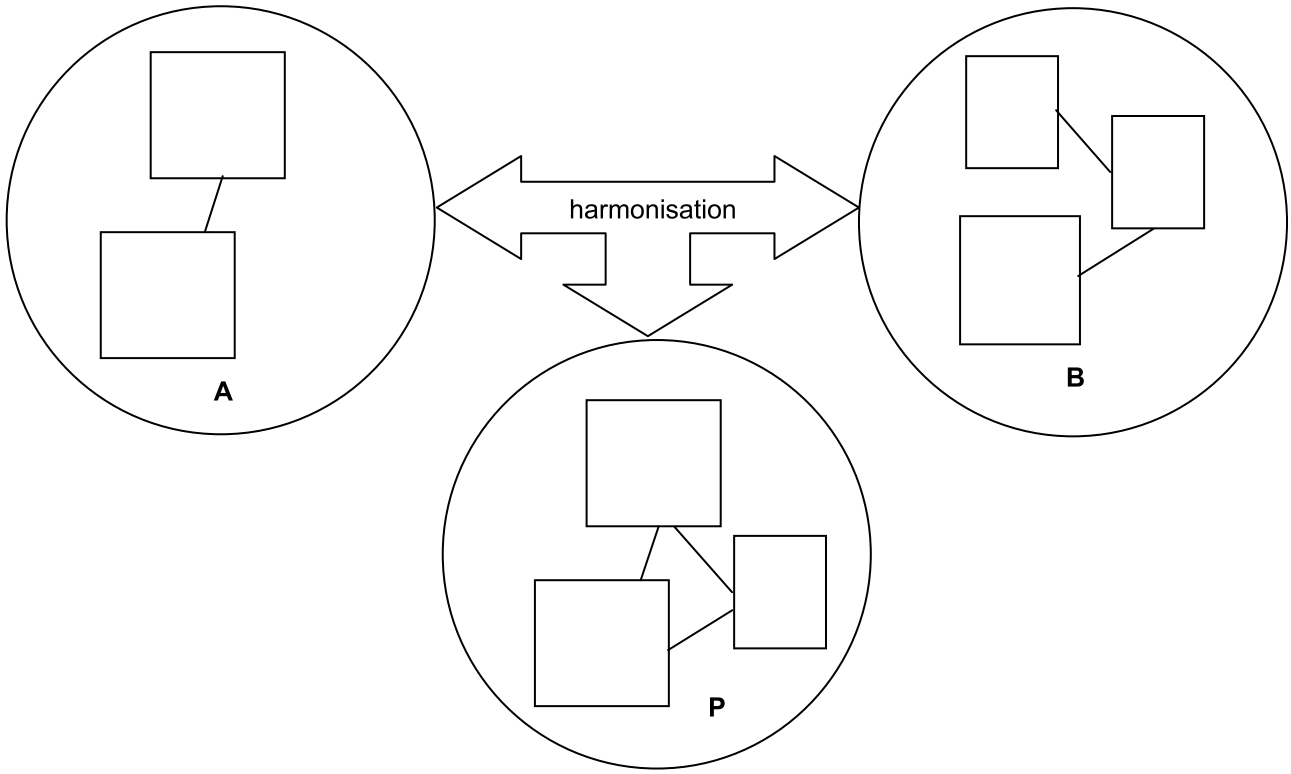


Figure 1— Data concepts from specifications A and B are harmonised producing preferred or reference set P

Figure 1 shows that data concepts from two semantically overlapping data specifications (labelled A and B in this example) can be harmonised to produce ‘P’, a preferred or reference set of concepts.

Harmonisation is typically more than a simple superset operation. If the original sets A and B have evolved independently then they may include differing versions of the same underlying concepts, and these are reconciled to a single preferred version in P. In some harmonisation contexts, some or all of the changes agreed for P will be fed back into the specifications A and B which become more interoperable.

The term "harmonisation" is used in a narrow sense to describe the process of selecting P by considering A and B, and also in a wider sense where it also includes any adjustment of A and B to achieve a better fit with P.

A and B in their original states may differ from each other and from P:

— for valid reasons of optimisation for their different business contexts;

— in cases where there is no strong justification for the difference.

The ideal goal of the overall harmonisation process is for all differences of the second kind to be removed, and for the only remaining differences between A/B and P to be clearly justified for the specific business context of A or B.

These process concepts and definitions may be understood more clearly by studying the specific examples in Clause 5.

4.4.3 Harmonisation and metadata registries

A harmonisation process is usually used with a metadata registry. The harmonisation process is used to increase the consistency of the registry contents. A and B are submissions to the registry, and the outcome P is also stored in the registry. In some contexts A and B will be replaced by P in the registry, while in other contexts A and B may be retained to support legacy systems.

A metadata registry imposes a structure (or "metamodel") for metadata. There are multiple registry standards and implementations with differences in metamodel. Most basic concepts of harmonisation are independent of the registry structure and could be instantiated with most registry implementations. However, some harmonisation rules require certain characteristics from the metamodel, and therefore place requirements on the type of metadata registry used.

A metadata registry may also impose a process for registration. The harmonisation processes described in this TR are typically an extra level of detail that is not covered within defined registration processes. They could be adapted to work together with most registry processes.

4.5 Steps in the harmonisation process

This subclause illustrates steps in the harmonisation process by continuing the simple example with the data specifications labelled ‘A’ and ‘B’, although a harmonisation process may involve more than two input data specifications, and may require more complex analysis at each step. Subclause 5.5 shows how these steps are instantiated in a specific harmonisation process.

In general for a pair of data concepts from (A, B) harmonisation is the selection of preferred concept *P* based on the attributes, relationships and semantics for individual data concepts in *A* and *B.*

P = h(A,B) where ***h*** is the harmonisation preference function

In some harmonisation contexts it may be possible to change concept A and/or concept B immediately to align with concept P. In other contexts, concept P is retained as a reference that should be used as the target for future developments.

The following steps must all be performed. **They are not independent** and are likely to be applied in multiple iterations.

4.5.1 Derive entity boundaries – including names and semantics

Simple illustration: *h(entity name A, entity name B)* $\Rightarrow$ *entity name P*

Harmonisation will identify a preferred or reference entity with a name. However, in generating the name, a process of 'conceptual normalisation' [Harmonise Project] is an intrinsic part of the harmonisation process. By agreeing a preferred name, a basis for the data concept is agreed at a highly abstracted level, without getting concerned with the structure of the concept. However, it should be noted that the subsequent harmonisation of detailed attributes and relationships puts a test on the results of the present step, as the full semantics of the entity include the semantics of its attributes and relationships. The harmonisation of attributes and relationships may cause a refinement to the harmonised entity boundaries.

4.5.2 Derive relationships – including role names and semantics

Simple illustration: *h(relationship A, relationship B)* $\Rightarrow$ *relationship P*

Relationships in A and B will lead to corresponding relationships between preferred or reference concepts in P.

4.5.3 Derive attributes – including names, semantics and datatypes

Simple illustration: *h(attribute A, attribute B)* $\Rightarrow$ *attribute P*

Attributes in A and B will lead to corresponding attributes of preferred or reference concept in P. Where two different attributes are semantically equivalent, a single preferred attribute must be chosen. In some cases attributes with related but different semantics may indicate the need for a new entity to hold the attributes.

4.6 Other work related to harmonisation

4.6.1 Extract, Transform and Load (ETL)

The idea of resolving differences in data sets covering related concepts but using different models is also encountered in enterprise data management, and is the focus of the "transform" step of the technique known as Extract, Transform and Load (ETL). There are a number of ETL frameworks which allow the specification of transformations from one data model to another. However the main focus of ETL is to establish a consistent data warehouse, not to harmonise models for future systems developments. Despite significant tool development we are not aware of a clear open standard approach to developing and specifying model transforms for ETL.

4.6.2 MOF QVT (Meta-Object Facility Queries Views and Transformations)

The Object Management Group have developed and published Meta-Object Facility (MOF) “Queries Views and Transformations” (QVT). It offers a graphical notation to extend UML, including transformation symbols. Like ETL, QVT can be used to resolve differences in models. However, it seems focussed on model-to-model transformation where metamodels are different, e.g. UML to relational database renditions of effectively the same domain model. It does not provide a process for harmonisation of similar but different data concepts.

5 Current approaches to harmonisation in ITS international standards

5.1 Four approaches

The four approaches to harmonisation in the ITS sector discovered in our investigation are described below:

5.1.1 ISO 14817 approach

TC204 WG1 developed ISO 14817:2002, which primarily specifies a data registry, but it also outlines a harmonisation process that uses the registry framework.

5.1.2 ISO/IEC 20943 approach

The ISO/IEC 20943 series reports on "Procedures for achieving metadata registry content consistency" and is being developed in six parts, two of which have been published. These reports expand on the ISO/IEC 11179 series of standards for metadata registries.

5.1.3 UN/CEFACT TBG17 approach

The approach developed within UN/CEFACT TBG17 is described in [TBG17 CCL (2008)]. The purpose of TBG17 is to be responsible for consistency and harmonisation of business process models and core components across business domains and sectors.

5.1.4 Highways Agency approach

This approach is taken by the Highways Agency (UK) to support its information exchanges with partner organisations.

5.2 ISO 14817 harmonisation

ISO 14817:2002 was based on an earlier edition of the ISO/IEC 11179 series, customising it for the domain of Intelligent Transport Systems.

ISO 14817:2002 describes its harmonisation process in A.6 in Annex A ITS/TICS Functional operating procedures (for the ITS data registry and data dictionaries).

ISO 14817:2002, A.6.1 Introduction states, *"These procedures detail how the CCC and the Stewards execute their responsibilities…regarding identification, reconciliation and documentation of data concept overlaps and duplications*…'.

ISO 14817:2002, A.6.2 describes a ten-step process for identification and resolution of harmonisation issues. The steps are summarised below.

5.2.1 Steps

Steps 1 to 4: The Registrar uses the capabilities of the Registry to identify potential overlapping or redundant data concepts. A listing of potential issues is prepared and distributed.

Step 5: Stewards analyse their assigned issues and determine appropriate resolutions.

'The first step in this process is for each of the Stewards involved in a set of data concepts at issue to understand the semantics of the data concepts at issue. If the semantics are not equivalent then the data concepts should remain separate. If they are equivalent or significantly equivalent, then the Stewards should agree to use one of them, modify one of them for joint use, or mutually agree to a new data concept to supersede those data concepts at issue. After achieving semantic resolutions, the Stewards then should address the Value Domain (if appropriate) of the data concepts at issue. The intent of this examination is to agree on a mutual solution to these dimensions of the data concepts at issue.

The Resolution may be that one data concept is selected and other data concepts reference the selected data concept as superseding, the data concepts at issue are merged into a new data element and the other data concepts at issue reference the new data concept as superseding, or the data concepts at issue are kept separate and independent.'

The Steward reports the resolutions to the Registrar.

Steps 6 and 7: Proposed harmonisations are distributed, with opportunity for further debate.

Steps 8 and 9: The Change Control Committee review and approve harmonisations and review issues.

Step 10: The Registry is updated.

5.2.2 Success of practical application of the process

An Australian prototype ITS data registry (supported by ITS Australia and other funding) operated for a brief period but achieved limited progress in data harmonisation.

The UK Highways Agency implemented an extended ISO 14817 registry. The registration process was found to be helpful, but not sufficient on its own to achieve harmonisation of multiple overlapping concepts, and a complementary process was defined as further described in 5.5 below.

5.3 ISO/IEC 20943 approach

The ISO/IEC 20943 series consists of six parts, of which two have been published (as Technical Reports):

— Part 1:Data elements;

— Part 3:Value domains.

Additionally, Part 5: Semantic metadata mapping procedure, is particularly relevant to harmonisation and has reached working draft stage.

All parts assume that the metadata adheres to the metamodel defined in ISO/IEC 11179-3.

5.3.1 Part 1:Data elements

This part describes two alternative approaches for populating a metadata registry: bottom-up and top-down.

The **"bottom-up approach"** involves harvesting data element definitions from system interfaces, and then inferring the abstract data element concepts. (The Highways Agency approach described in 5.5 can also be described as "bottom-up").

To quote the most relevant section of Part 1:

"…the registrar should perform content research to determine the following:

• Is the data element described in an existing International, National, or organizational standard?

• Does a data element exist in the registry, or a federation of registries, that has the potential for being reused?

The practitioner will determine if a data element might be adapted to meet new requirements, or if some attributes of an existing data element (e.g., value domain, data element concept, or conceptual domain) might be reused with the new data element. Content research should include a search of conceptual domains, data element concepts, and value domains, as well as data elements, to identify attributes that might be relevant to the data element to be registered. If a standard data element exists that can be used as a model to meet the particular specifications for a new purpose, some of its associated metadata items may be reused for registration of the new data element.

The result of this step is confirmation that a new data element is needed, or a decision to modify or reuse an existing data element or some of its attributes."

After the data elements are well described, the abstract data element concepts are derived. A single data element concept may be related to several data elements with different representations.

The "**top-down approach**" is a planned approach to defining metadata, as opposed to being in response to an existing system interface. In ISO/IEC 20943-1 the top-down approach proceeds in an object-oriented way, registering object classes before specific data element concepts, and data element concepts before data elements and representations.

There should be little need for harmonisation because the top-down registration should not create any competing concepts.

5.3.2 Part 3:Value domains

Part 3 allows for the registration of similar but different value domains. When considering registration of a new value domain, the registrar should determine whether any existing conceptual domain is appropriate, and if so, the sets of value meanings should be checked and any new value meanings added. There is no guidance for what to do if a pair of value meanings is similar but different.

Part 3 also allows the approach where two conceptual domains remain separate but each relate to a new common super-ordinate conceptual domain with no enumerated values.

5.3.3 Part 5:Semantic metadata mapping procedure

This part is in progress at the time of writing. Working drafts describe how to map between overlapping metadata sets describing similar concepts. The mapping is performed specifically for the purpose of supporting the identification of a recommended set of data elements which may be a new combination of the existing metadata sets.

A three-stage process is outlined:

1. Identify the metadata sets that will be in scope for mapping.

2. Group data elements by object class. An approach noted as helpful is to take one metadata set as "primary" and to follow its groupings. For each object class, list the corresponding elements within each metadata set.

3. Find common domain element concepts within each object group. Identify the "type of heterogeneity" that exists for each specific data element. Derive a set of recommended data elements from the candidates.

There are six categories of heterogeneity:

— Same, no difference;

— Hierarchical difference;

— Domain difference;

— Lexical difference;

— Syntactic difference;

— Complicated difference;

Some categories have further sub-types.

Part 3 also recognises that value domains must be harmonised and identifies three kinds of conversions and three kinds of structural rearrangement.

Although Part 3 gives a process to identify sets of corresponding data elements, and identifies the kinds of difference between these elements, it does not include any guidance on how to choose between the competing alternatives once the differences have been identified.

5.4 TBG17 Business process & core components

5.4.1 Fundamental idea of core components

UN/CEFACT has published the “Core Components Technical Specification”, part of the ebXML Framework. The key idea that supports harmonisation is the separation of “core components”, which have no specific business context, from “business information entities”, which apply in specific business contexts. Core components can therefore be viewed as an example of the background ontology described in 4.4 of this Technical Report.

Core components have some similarity to the idea of "data element concepts" in ISO/IEC 11179-3: both are abstractions away from specific business contexts. However, core components are still closer to implementations than data element concepts; for example core components have datatypes and can have structure similar to that used in implementations.

5.4.2 Harmonisation team submission guidelines and procedures

The Core Components Technical Specification does not provide fully comprehensive guidance for the harmonisation process. Recognising this, UN/CEFACT TBG 17 developed “Core Components Library Submission Guidelines and Procedures” to guide the maintenance of its "Core Components Library".

A normal submission to TBG17 consists of core components and business information entities, plus a supporting business requirement specification. TBG17 then harmonises each set among domains to ensure that the harmonised set may be used across different business process domains.

Important steps in the process are:

- To ensure that submissions are well-formed, using a checklist;
- To harmonise core components;
- To edit business information entities based on harmonised core components;
- To harmonise business information entities.

Submitters must supply all details in a detailed template form, which appears to prevent submissions consisting solely of existing electronic models or schema.

Figure 2, reproduced from the TBG17 guidelines, illustrates the process of harmonisation of core components.

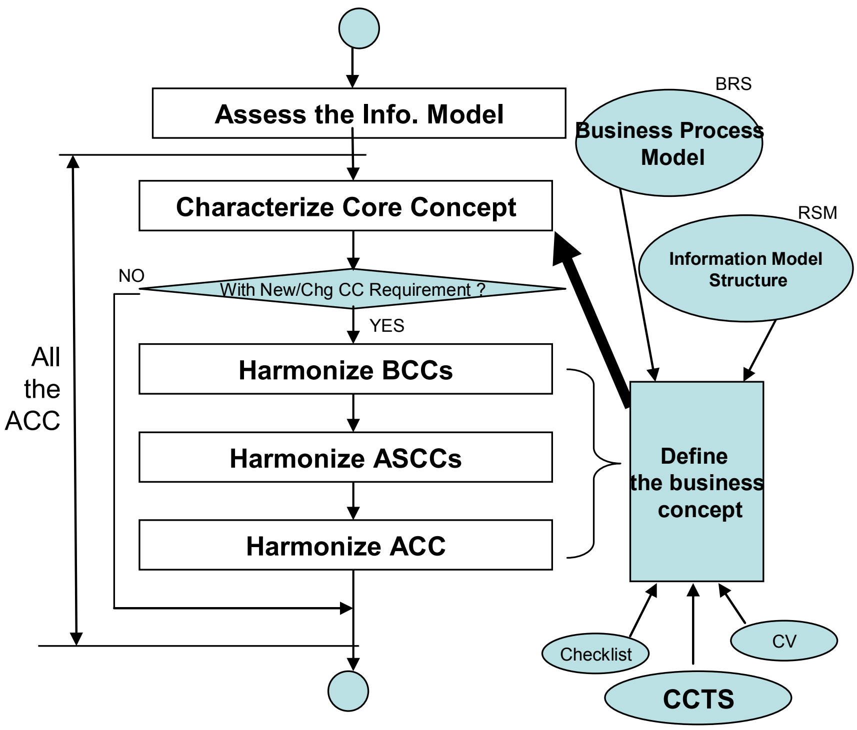


Figure 2 — TBG17 Core Component Harmonisation

A similar process exists for harmonisation of business information entities.

The TBG17 Core Components Library has a defined version control process, with a release going through four stages as contents go through the harmonisation process.

5.4.3 Harmonisation rules

To guide the individual steps of these processes, the submission procedures document contains a section on "Harmonization Rules and Guidelines", items which TBG17 has learned through the experience of Harmonization. Many of these rules are concerned with metamodel conventions, adding extra constraints to the general metamodel of the Core Components Technical Specification, but two rules in particular give guidance for the actual harmonisation part of the process:

— "The submission SHOULD NOT include BCCs [basic core components] which are generic in nature where they are not backed by a business requirement (submission) and where they have no clear business semantics. Based on the business semantic meaning, BCCs SHOULD be named as generic as possible. Note that business domain qualification occurs at the BIE level in context."

— "The Oxford English Dictionary in conjunction with the TBG17 Controlled Vocabulary SHOULD be used as reference source for meanings of words in definitions and DEN (Dictionary Entry Name) Terms."

An earlier (2004) version of the same guidelines gave additional specific and interesting guidance and rules on the subject of harmonisation, but these appear to have been withdrawn:

— "Harmonisation should ensure that a single semantic concept is captured in one and only one Core Component structure. This may conflict with different views on that concept in different contexts and it conflicts with the emergence of new submissions."

— "If it is agreed, that there is a need to have two or more core component structures for the same semantic concept, then the core components library administration has to make sure that they refer to each other in order to guarantee that any further development will be a consistent one."

The earlier version also included a concept of simplified "derived" core components that referred to more complex core components; this concept appears to have been dropped.

5.4.4 Success of practical application of the process

TBG17 has established a library of around 2000 core components and business information entities, although most have not required harmonisation due to being unique.

5.5 Highways Agency (UK) – Core Components Analysis of the ITS Metadata Registry

5.5.1 Basis

The UK Highways Agency has also derived a process of Core Components Analysis to encourage harmonisation of data concepts. The process is based on the idea from the UN/CEFACT Core Components specification that “core components” have no specific business context, while “business information entities” apply in specific business contexts. The process was initially developed independently from the UN/CEFACT TBG 17 guidance, but produced many similarities to that guidance. However, it differs in scope and in detail from the process applied by UN/CEFACT. The UN/CEFACT process aims to ensure global interoperability. The Highways Agency approach is more focussed on incremental improvements to legacy systems and specifications.

The process uses the extended ISO 14817 metadata registry implementation (which can be viewed at www.itsregistry.org.uk, with the relationship to ISO 14817 explained at www.itsregistry.org.uk/14817.html). The submitted data definitions all have their individual business contexts and they can be seen as context-specific instantiations of a single set of underlying concepts, the core components, as shown in the example of Figure 3.

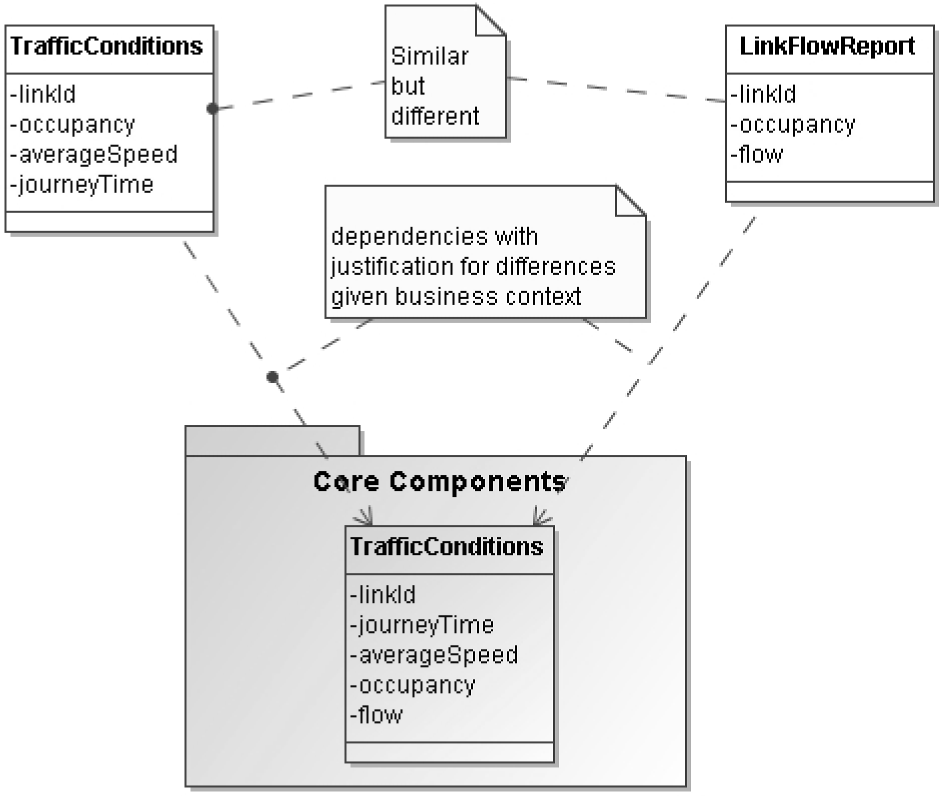


Figure 3 — Core component example

5.5.2 Derivation of core components

The set of core components is derived through analysis of the existing registered models. The process aims to be as objective as possible to avoid the possibility of the core components being yet another competing model. New core components are not introduced unless justified by existing models. The engineer should resist the temptation of “fixing” the registered models unless there are other registered models or established design rules to justify the improvement. Therefore, while UN/CEFACT Core Components are supposed to be independent of business context, the core components of the Highways Agency’s metadata registry do have a broad business context that is the superset of all the business contexts of the submitted models. This helps ensure that any analysis effort is constrained to be focussed on Highways Agency business needs. For this reason the core components analysis process is also referred to as "simple ontology", because it is analogous to the agile programming practice of "simple design", where the design is only extended in response to incremental current requirements and not future requirements.

The outputs of the analysis process are the set of core components plus the mappings from registered models to core components. Each link observed between a registered item and a core component has associated text attempting to explain and justify the differences. If there are legitimate justifications for the shape of the model, those will be declared, but if the model contains poor design or sloppy thinking, these should also be exposed because a good business justification cannot be agreed. This gives a powerful way to provide objective review feedback and improve the quality of submitted models, since there is often no genuine justification and the submitter can understand and agree this. Furthermore, the improvements made through this process are in line with requirements brought by other models, and any changes should bring the submission closer to the others.

5.5.3 Contrast with UN/CEFACT process

The Highways Agency process is therefore slightly different from the UN/CEFACT process (depicted in Figure 4), which starts with business requirements and then derives the core components and specific business information entities. The business requirements may or may not have been derived from existing systems.

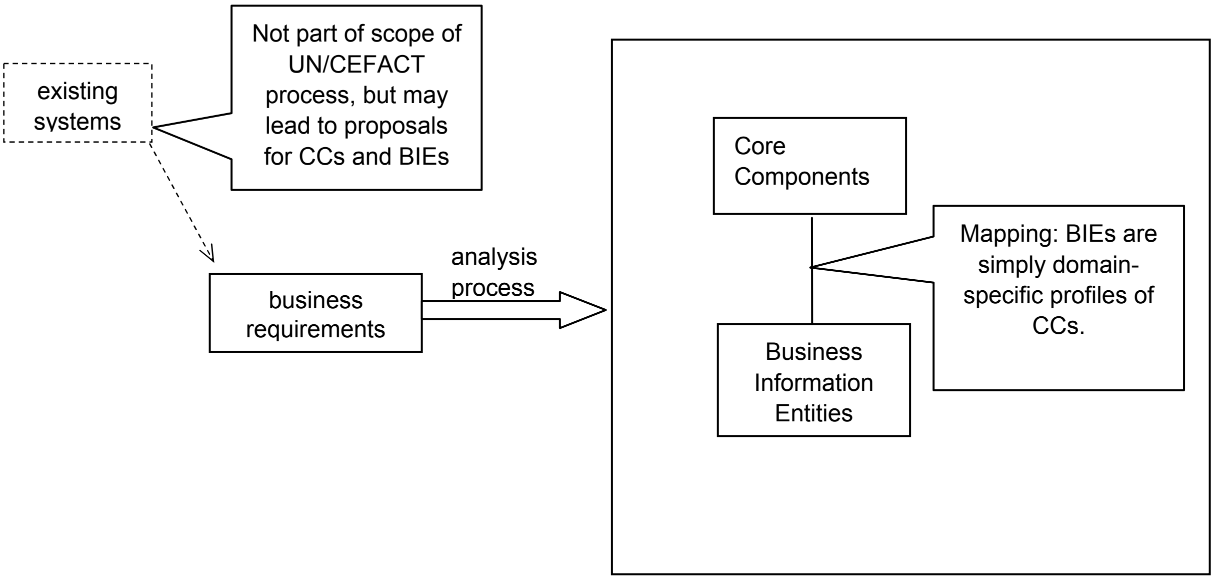


Figure 4 — Core components analysis process of UN/CEFACT

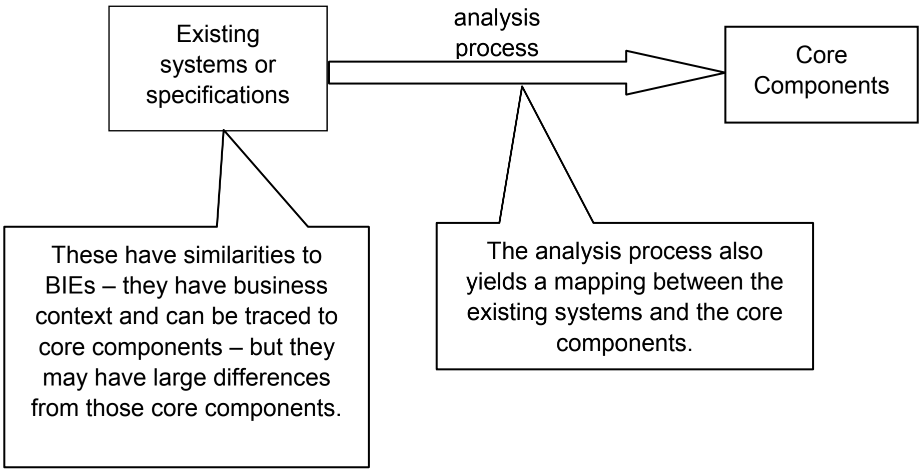


Figure 5 — Core components analysis process of Highways Agency

In the UN/CEFACT approach, the “Business Information Entities” may be created at the same time as the core components and may differ from the existing legacy systems.

In the Highways Agency approach (depicted in Figure 5) the existing data specifications are expressed alongside the core components and shown to be context-specific instantiations of those components, but because the existing specifications were designed before the core components, the mappings allowed between them to have to be much more flexible than the mappings allowed by the UN/CEFACT Core Components metamodel.

In the UN/CEFACT approach, the newly derived BIEs provide an explicit target for the future evolution of legacy systems. In the Highways Agency approach, this target is not so explicit, but it is suggested by the combination of the core components and the words of the mappings. In the Highways Agency context, the business case for the effort in creating explicit *ideal* BIEs is not clear as the stakeholders for the submitted specifications may not be receptive. Instead, the process looks case by case on the feedback that can be given to stakeholders of existing systems. A compensating advantage of the process is that the mappings from legacy systems to core components are explicit.

As the Highways Agency core components are created in response to multiple independently developed systems or specifications, perhaps containing flawed design, the differences between those systems and the core components are greater than those allowed by the UN/CEFACT Core Components metamodel, so the mapping has to be more flexible. The mappings are expressed as UML dependencies, and each model element from existing specifications may have one or more dependencies to any model elements in the core components model.

5.5.4 Rules guiding the analysis

The core components analysis is performed for a given subject matter area where harmonisation in that area is of particular business benefit to the Highways Agency. All registered definitions in the subject matter area are considered. The main steps are similar to those defined in 4.5.

1. Build up an outline conceptual schema by including concepts one by one from individual registered data definitions. Show how each submitted entity maps to the core component entities.

2. Fill in the attributes of the core components by considering each submitted attribute. Show how each submitted attribute maps to its corresponding core component. This may lead to an alteration of the entity boundaries from the first step.

The core components provide background ontology. For that purpose alone, the core components could be expressed in an abstract way. However, to encourage harmonisation it is desirable for the constructs used in core components to have a similar level of abstraction to the constructs used in submitted models. The core component attributes, therefore, use specific concrete datatypes, and the analysis process follows a guiding rule:

— If the core components were implemented, they should be able to represent any data currently representable by the submitted models (within the chosen scope areas of the core components) with the added condition that it would be permissible to have empty attributes and associations in the core components implementation.

If the submitted models are similar in approach, then it can be relatively straightforward to create core components. If the subject matter area involves a complex taxonomy, then it may be worthwhile defining how the submitted items map to a conceptual taxonomy before proceeding to the full conceptual schema. If the submitted models take radically different approaches because of justified business requirements, it can be difficult to objectively choose the best way to represent the core components.

The following rules supplement the guiding rule above in the selection of core components. It is interesting that UN/CEFACT TBG 17 also evolved very similar rules independently.

— Choose the most general mechanism that preserves all the semantics of the original structure.

— While a specific model may allow the representation of a single concept in two ways, for example to allow backwards compatibility, the core components should not give the choice - as long as an implementation of a single mechanism would be capable of conveying the same semantics.

— All names must follow a stated detailed style policy. When submitted styles differ, the choices of core component names are therefore made objectively.

Core components analysis can be time consuming, and in particular the registration of the mappings can produce high maintenance if the registered models are not stable. The approach of the Highways Agency project is to proceed with core components analysis only in subject matter areas where further modelling or translation work of relevance to the Highways Agency is imminent or on going. Mappings for submitted models are registered only when those submitted models are relatively stable.

5.5.5 Portraying mappings between entities

The Highways Agency approach has developed similar ways of portraying the mappings to those developed by TBG 17. Figure 5 shows an example of the mapping of entities and relationships, and Figure 6 shows an example of the mapping of attributes. Occasionally, where there is a complex taxonomy in the subject domain, it can be useful to analyse the taxonomy before proceeding to the full conceptual schema. Figure 7 shows an example of taxonomy mapping. In all these figures the core components are the shaded elements shown as targets of the dependency arrows, while the submitted models are the clear white elements from which the dependency arrows originate.

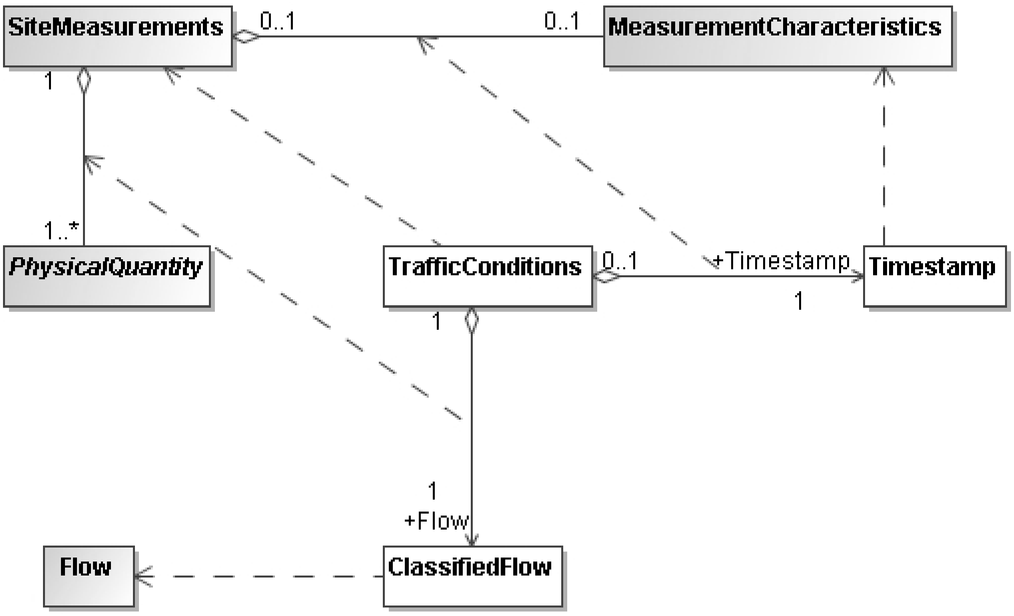


Figure 6 — Mapping of entities and relationships

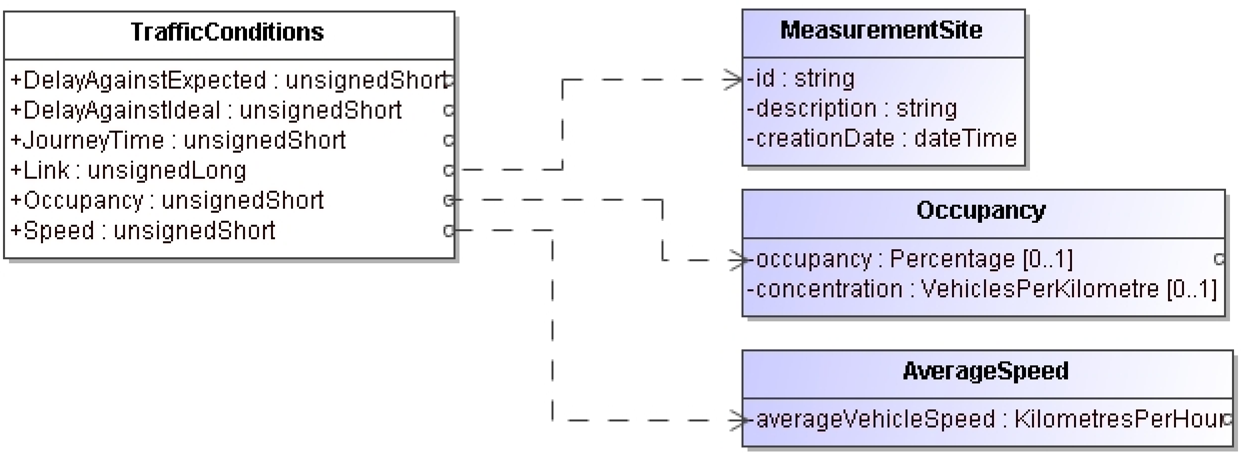


Figure 7 — Mapping of attributes (fragment)

In Figures 6 and 7, a relatively simple model has been harmonised with other models with greater scope. At least one of those other models used the term "averageVehicleSpeed" where the simple model used the term "Speed". The intended semantics were actually the same, so the submitted "Speed" attribute is mapped to the core component attribute "averageVehicleSpeed". The other differences are also due richer or more accurate or precise modelling in the other models.

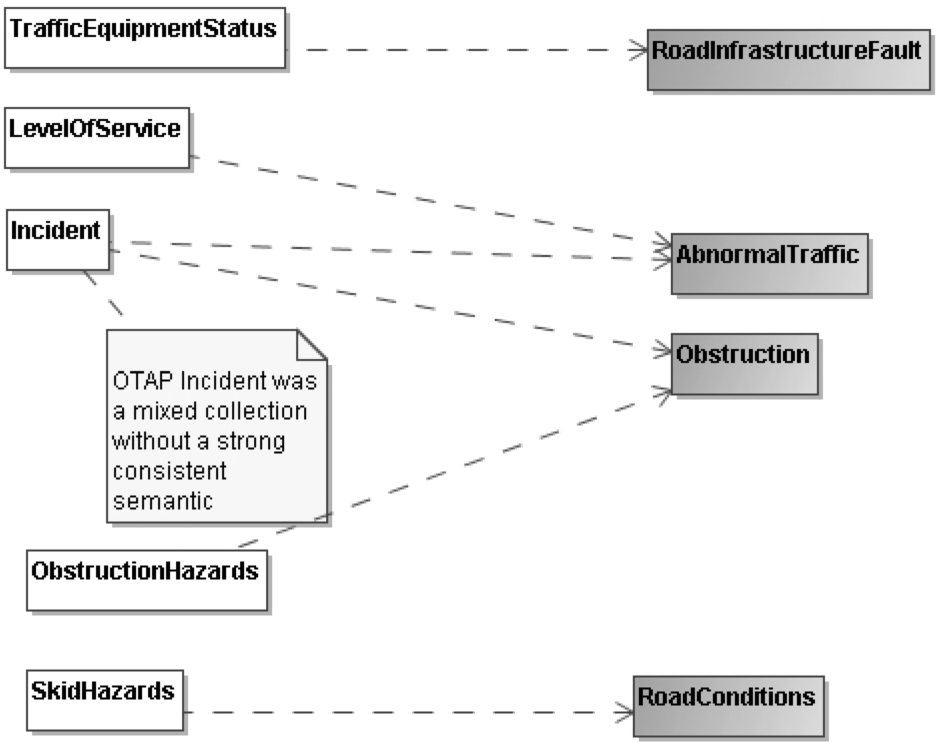


Figure 8 — Mapping of taxonomy (fragment)

Classification schemes (taxonomy) used for traffic events varies widely across different submitted specifications. Submitted models typically use some kind of abstract traffic event class with specific subclasses, and the same pattern is reflected in the core components. A single generic and consistent taxonomy was chosen which could represent all cases, and the mapping from each submitted model was explained using diagrams such as Figure 8.

Each dependency is given documentation (available in the registry) that states the business context, explains the differences to the core components, and if possible justifies these differences.

5.5.6 Precise mappings

When a submitted model element is simply a context-specific instantiation of a core component (which is always the case in the UN/CEFACT TBG 17 approach), there is a clear and simple one-to-one relationship between concepts and the mapping needs no further specification. For example as Figure 9 shows, an instance of the specific registered class AxleSpacing with two values corresponds to an instance of the Core Component AxleSpacing with identical values (the mapping of datatypes being specified elsewhere).

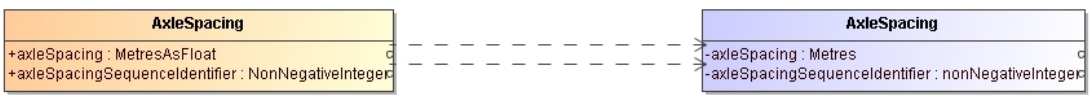


Figure 9 — Simple mappings need no further explanation

However, as explained above, the Highways Agency core components are created in response to multiple independently developed systems or specifications, and the differences between those systems and the core components are greater than those allowed by the UN/CEFACT Core Components metamodel, so the mapping has to be more flexible.

At their simplest the mappings are simply pointers to related concepts. At their most precise the mappings could carry sufficient detail to completely specify a translation from one format to another.

The Highways Agency has developed a supplementary notation that allows core component mappings to be sufficiently precise and unambiguous that they specify a data translation. The types of translation operations identified go beyond those identified in ISO/IEC 20943-5.

Recall that each UML class and attribute already has a UML dependency from the specific registered model to the core components model. To specify the transformation of values, we can label the dependency with a statement constructed of up to four clauses:

- from;
- where;
- to;
- with.

Figure 10 shows examples of each of these clauses. The full mapping is specified in Annex A.

The precise clauses remove ambiguities that would otherwise be present in the mappings. As a valuable side-effect, the rigour required to create the mappings often leads to improvements in the core component models. Although the core component derivation process is as objective as possible, it is easy to make a mistake unless the derivation process is rigorous.

The disadvantage is a high maintenance cost. If the core component mappings for a submitted model are registered, and either that model or the core components change, the mapping must be updated or removed. The Highways Agency therefore decided to derive and register precise mappings only in the most important cases.

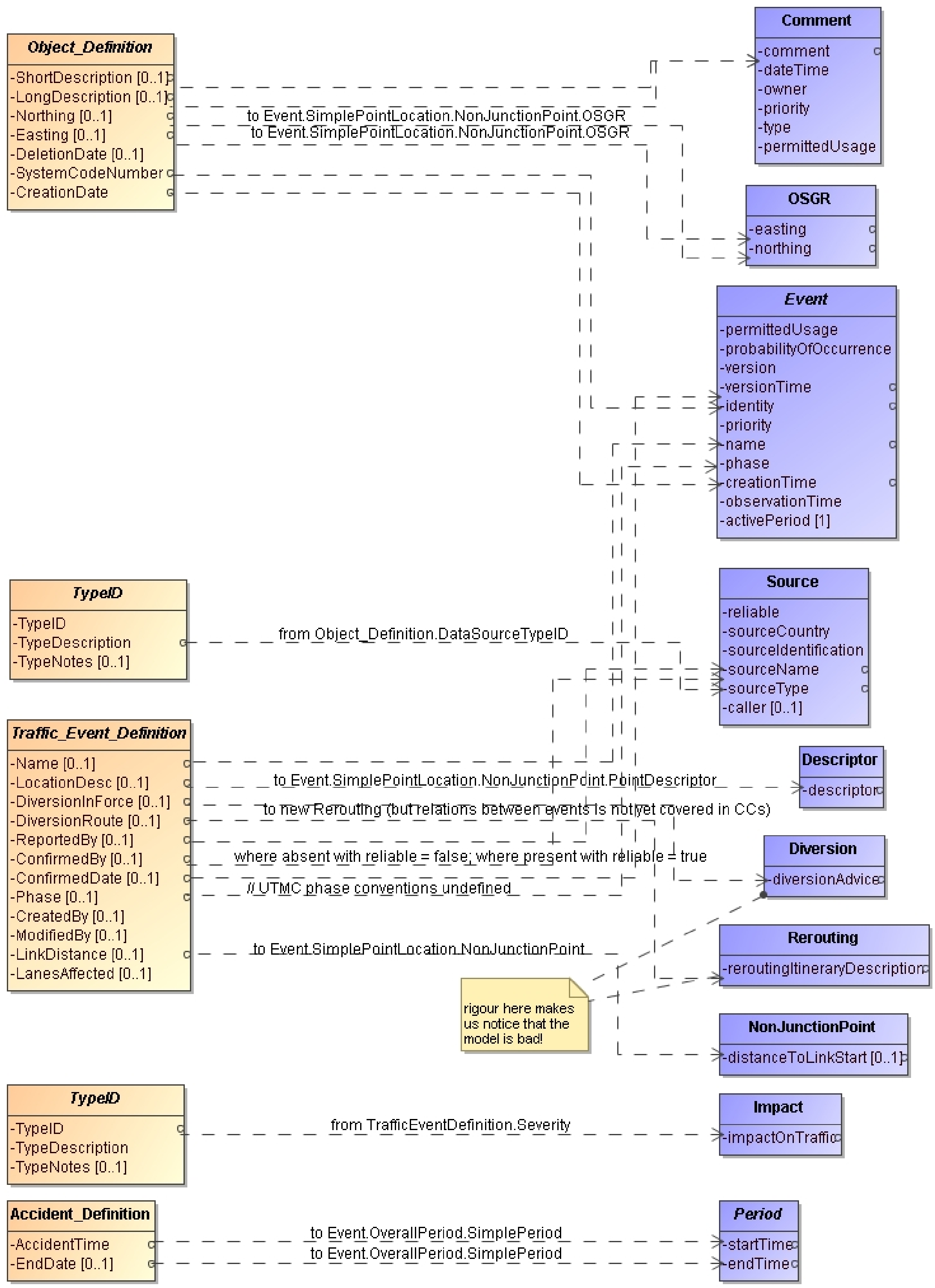


Figure 10 — Example of precise mappings

5.5.7 Success of practical application of the process

The only instantiation of this process is in the UK Highways Agency's ITS Metadata Registry.

At the time of writing there are a few hundred core components out of a registry population of over 20,000 model elements. Each core component subject area has been derived from mapping at least two overlapping submissions, with up to eight submissions in the area of traffic events.

The process of derivation of core components and mappings has produced feedback to submitters to improve quality, and in some cases there have been resubmissions with improved quality. The resulting models have increased value and reduced costs for their users. However, registry feedback is not always acted upon by submitters because as yet the registry has no official mandate to enforce quality levels.

The core components analysis has made it clear how an element in one model is related to a semantically similar element in another model. This understanding can be used to develop translators and has already been used to develop a translator between the European-funded DATEX II specification and UTMC, the UK urban traffic management standard.

The registry has provided a focus for activity on convergence of system interfaces and data standards. It has been used by working groups dealing with data standards and technology convergence. However, progress on convergence directly attributable to the registry is restricted by the registry not yet having an official mandate to be able to enforce recommendations.

6 Harmonisation approach for designers of data specifications

The processes and guidance defined in the four harmonisation approaches described in this report are intended for use by a central registration authority or harmonising committee, responsible for harmonisation across multiple specifications, rather than for the writer of a single data specification. This clause provides guidance for the writer of a new data specification.

When designing a new data specification, the designer should seek relevant existing data definitions in metadata registries and in published standards and specifications.

Even if there are no existing relevant application-level definitions, the designer should still seek to re-use the following from published standards and from existing metadata registries, rather than inventing new ones:

- Naming conventions (for example ISO/IEC 11179-5 Naming and identification principles);
- Datatypes [for example ISO/IEC 11404 "General-Purpose Datatypes" (GPD)];
- Data definition and data modelling practice and style rules and guidelines (for example ISO/IEC 11179-4 Formulation of data definitions, and ISO/IEC 20943-1).

If however there is some overlap with existing data definitions, and yet a legitimate business reason to develop new definitions, the designer should seek existing core components in the application area.

6.1 Where there are relevant core components in a metadata registry or library

The designer should:

a) Explore any registered mappings from the relevant core components to specific registered data specifications.

b) Consider element-by-element whether the existing data specifications are suitable for the new purpose.

  - If so, re-use the element in the new data specification.
  - If not, create a new context-specific instantiation of the core component, tailored for the new purpose.

c) Submit the resulting data specification to the metadata registry.

In some cases the core component may be conceptually similar to what is required for the new purpose, but not exactly correct. In this case, the designer should build the necessary characteristics into the new specification and when this is submitted for registration the registrar should generalise or re-factor the core components to cover the new characteristics.

6.2 Where there is no relevant registry using core components

Ideally the designer can still evolve the data specification following the principles of core components. When a designer considers an existing data concept for re-use, but decides that it is not quite right for the new requirements, this may be because both are specific instantiations of a single domain concept where each business context brings slightly different requirements.

Ideally for the ITS community, the designer can make these core components explicit.

For example in Figure 11, there is a requirement for a data element (in core components terminology a business information entity) which is similar, but not identical to the existing data element A. The designer has a new element B in mind, but recognises that both A and B are context-specific versions of a single concept C.

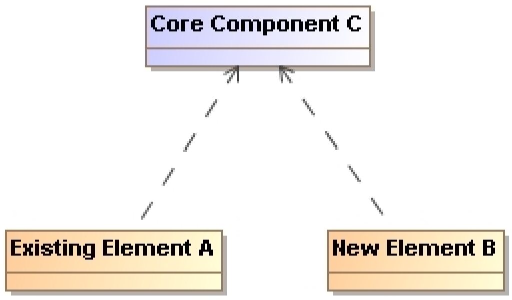


Figure 11 — Deriving new data elements and core components

The core component should be a simple generalisation that describes the underlying domain semantics of the associated data elements. The new element should be the simplest specialisation of the core component for the new business context. Its semantics and syntax may even be identical to those of the core component.

There may be benefit in making the core components known to the maintainers of the original data specification. They may see some ways to improve or generalise their specification in a future version in order to harmonise with the new specification.

If the core components can be published and maintained in a metadata registry, the thinking of the designer is made available to potential users and future designers who can then understand the relation of the specifications, how they can interoperate, and how to achieve harmonisation with further new specifications.

7 Harmonisation as a means to improve efficiency

In WG1's research of many initiatives around the world to tackle the problem of inconsistent data representation in order to develop this Technical Report, it was interesting to see that in many of the projects the resultant costs of inconsistent/incompatible data representation were considered so obvious that the benefits were assumed rather than clearly stated. And indeed they are very obvious and can be summarised in just a few points:

Inconsistent data definition (semantic difference in use of a name) causes

— Data to be misinterpreted (with the consequent and potentially life threatening bad decisions taken on the basis of incorrect information).

Inconsistent data representation causes:

— Data to be unusable (with the resultant costs for finding / capturing alternative data).

— Data to be misinterpreted (with the consequent and potentially life threatening bad decisions taken on the basis of incorrect information).

- Data to be reformatted where a conversion can be reliably made.
- Lack of interoperability of data concepts.
- Lack of mobility of data across services (and the consequent limitation in the potential provision of such services).
- (significant) duplication and repetition of the cost of developing and defining data concepts for the same semantic objective and also the consequent time delays incurred.

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. The net result of data harmonisation can be summarised as follows:

- Services are delivered on time and as agreed.
- Data is communicated effectively (unambiguously) to the service.
- Data is unambiguously comprehensible to the receiver.
- Data may be available to other not directly related services to improve their performance, or indeed even enable them to occur at all.
- Data concepts can be stored and addressed using common data registries and data dictionaries.
- Avoidance of the costs and time delays incurred by the duplication of developing and defining multiple data concepts with the same semantic objective.

8 Conclusions and recommendations

The business value of the harmonisation of data concepts rests in the provision of services that are delivered on time and as agreed; the efficient and unambiguous communication of data; the provision of data that is unambiguously comprehensible to the receiver; the availability of data to other indirectly related services to improve their performance, or indeed even enable them to occur at all; data concept storage and addressing using common data registries and data dictionaries; and avoidance of the costs and time delays incurred by the duplication of developing and defining multiple data concepts with the same semantic objective.

The harmonisation process has to satisfy semantics as well as structure. The harmonisation process must:

- Agree the form in which the data in general will be defined in standards, data registries and data dictionaries (normally ASN.1 data definition modules in respect of ISO TC204); in other words agree the metamodel.
- Agree the meaning of each data concept that is intended to be harmonised and give it a name.
- Agree the structure and content of the data concept.
- Agree the format(s) in which the data will be encoded (XML, EBXML, human readable, binary, etc.).

This report has compared four approaches: The ISO 14817 approach (built on ISO/IEC 11179), the ISO/IEC 20943 approach (also built on ISO/IEC 11179), UN/CEFACT TBG17, and the Highways Agency approach. Each brings its strengths to data harmonisation. We recommend the strengths of an underlying ISO 14817 or ISO/IEC 11179 metadata registry. ISO/IEC 20943 usefully enhances the ISO/IEC 11179 specification, but may require additional measures to deal systematically with competing concepts with complicated differences. The two most highly developed and practically used processes are both based on the idea of UN/CEFACT Core Components and this report therefore recommends the use of Core Components for harmonisation.

ISO/IEC 11179 data element concepts provide a higher level of abstraction than core components. In general, people should work with as much abstraction as they require, and no more. Working at a high level of remoteness from specific business purposes risks the work not being useful for any business purpose. However, there are often good reasons to use high levels of abstraction, e.g. to allow re-use of higher level specifications across multiple specific contexts, possibly with integration benefits. Where the ISO/IEC 11179-3 metamodel is being closely followed, ISO/IEC 20943 advice specific to that metamodel should be followed.

The choice between the Highways Agency and TBG 17 processes for core components should be made by considering how their focus fits the context and scope of the harmonisation.

Use the Core Components Analysis process developed by the Highways Agency when:

— Specific legacy systems and legacy data standards are important.

— Existing systems will progress only through small incremental improvements.

— It is important to understand the similarities and differences between different existing system interfaces and data standards.

— The harmonisation team does not have a strong mandate for change to existing systems. Some limited harmonisation can still be achieved.

Use the UN/CEFACT TBG17 process when:

— Designing new data standards for the future.

— Legacy systems and legacy data standards are not a focus.

Annex A
(informative)

Conventions for precise core component mappings

This annex presents the detailed conventions for precise core component mappings described in the main text.

A.1 Conventions for mapping statements

Recall that each UML class, attribute and association in a specific registered model has a UML dependency to an element in the core components model.

To specify the transformation precisely, we can label the dependency with a statement constructed of up to four clauses:

— from;

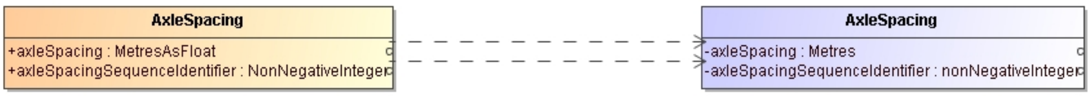
— where;

— to;

— with.

A.1.1 Empty statement

In simple cases the translation can be understood from the dependency without further annotation. For example, an instance of the registered class AxleSpacing with two values becomes an instance of the Core Component AxleSpacing with identical values (the mapping of datatypes being specified elsewhere).

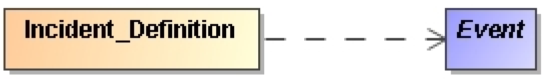


In all examples in this Annex the specific model element is shown on the left while the core component is shown on the right.

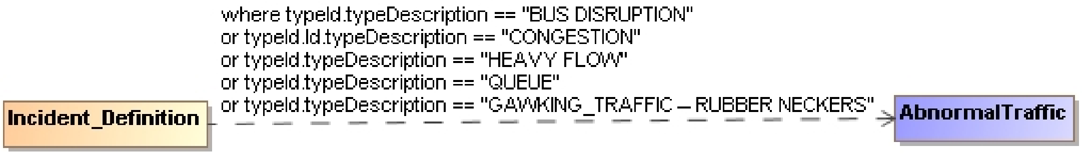
A.1.2 “where” clause

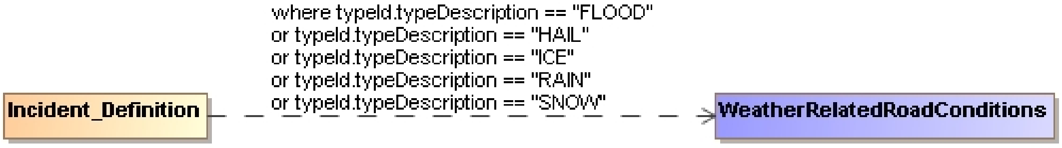
If the specific registered model element is more general than the core component, a “where” clause can select the subset of values for which the given mapping applies.

For example the registered class “Incident_Definition” can represent a number of distinct concepts such as abnormal loads, obstructions, hazards and weather. Without precision, this forces a mapping to the top level traffic event core component.



However, this mapping says very little and fails to convey the meaning of Incident_Definition. Instead the Incident_Definition can map to specific core components with the use of a “where” clause.

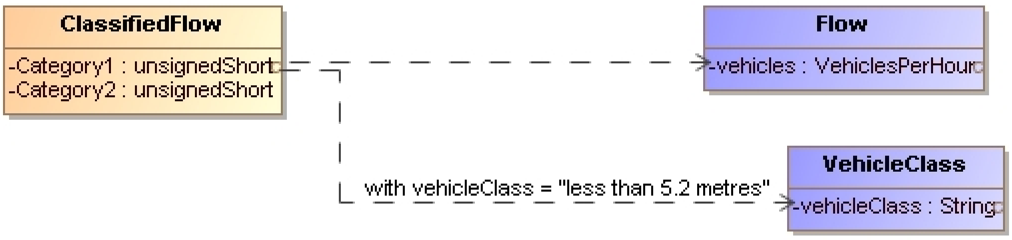




The illustration shows just two examples, and there would be more in the same style to completely cover the semantics of Incident_Definition. This is a fairly extreme example – in many cases a simple single "where" clause is sufficient.

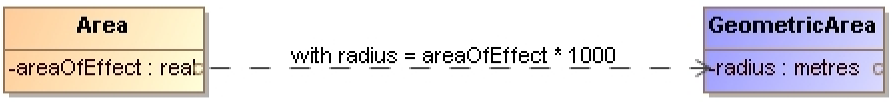
A.1.3 “with” clause

A “where” clause deals with a specific registered model element being more general than its corresponding core component. It is also possible that a specific registered model could be more *specific* than its corresponding core component element. For example a specific registered model could have used a very specific class to represent a concept that is expressed by a particular enumeration value within a more general class in the core component. The transformation can be expressed by a “with” clause.

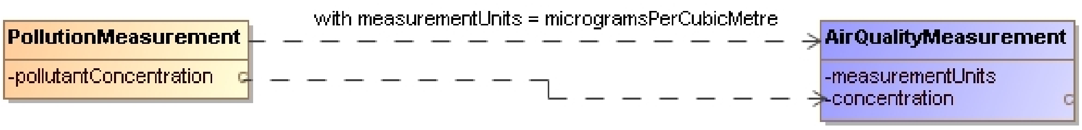


In this example the attribute “Category1” has two dependencies and therefore produces two attribute values in the core components: the numeric value is mapped directly to Flow.vehicles and there is an accompanying VehicleClass instance with the specified value of vehicleClass attribute, which comes from the definition of “Category1”.

A “with” clause can also be used to express value mappings from attribute to attribute where the values do not use the same value domains or the same units. For example when converting an attribute from kilometres to metres, the following clause would be used:

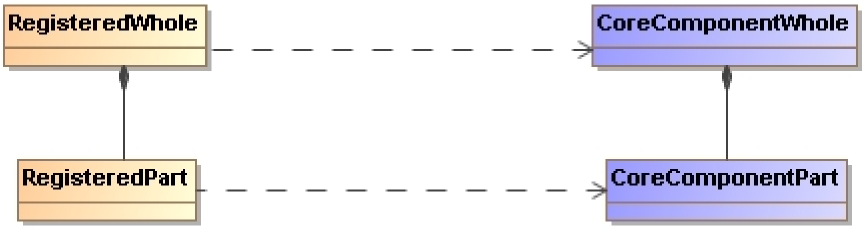


A further example is where units are constant in the specific model but variable in the core components.



A.1.4 “to” clause

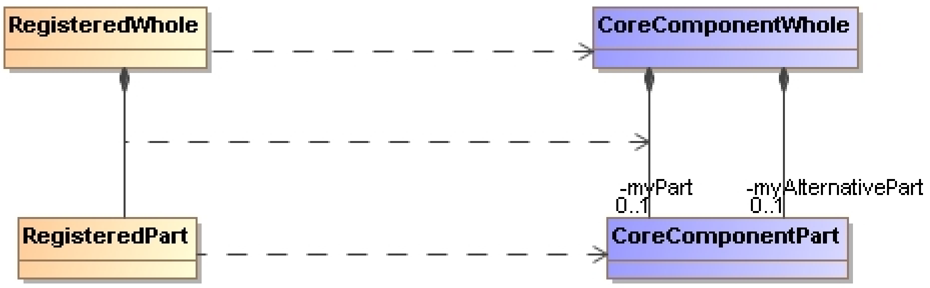
A translation will often deal with a set of related classes at the same time. The links between the instances must be preserved. In many cases an unlabelled mapping is sufficiently clear. In the example below, it is understood that the RegisteredPart instance belonging to the RegisteredWhole will map to a CoreComponentPart instance belonging to a corresponding CoreComponentWhole. If necessary a systematic rule could be expressed to formalise this understanding.



In other cases these simple mappings could be ambiguous. In the following variant the CoreComponentWhole can now have two different kinds of association to CoreComponentPart, with different semantics.

The plain mapping of classes is now ambiguous. In order to represent semantics, and in order to support a translation, we need to specify which kind of CoreComponentPart is the equivalent to a RegisteredPart when used in the context of a RegisteredWhole.

The principal way to remove ambiguity is to map the associations. By adding a dependency we specify which association is the equivalent to the registered association, and therefore which role the RegisteredPart is taking.



An alternative notation that is convenient on attribute mapping diagrams is to use a "to" clause.

The "to" clause becomes especially convenient when a registered class is used in multiple contexts and when the mapping of its attributes differs between the contexts. A single diagram showing all of this with associations would be very cluttered.

A.1.5 “from” clause

In the similar but reversed scenario, a single model element may be used in multiple contexts within an aggregating whole within the specific registered model. For example the registered class Incident uses the class TypeID to describe the type of incident and also the type of data source. The mapping shown below is ambiguous: does it apply to the TypeID of the Incident_TypeID or the TypeID of the Incident’s DataSource_TypeID?

Again, the ambiguity could be removed by showing the association mappings on the same diagram, but a convenient alternative is to use a “from” clause.

A.1.6 Combining clauses in statements

It is possible to have any, all, or none of the clauses in each statement. The same mapping may also have multiple clauses of the same type.

For example a single dependency could have two sets of four clauses:

from AClass.assocRole *where* anAttr == “value 1”

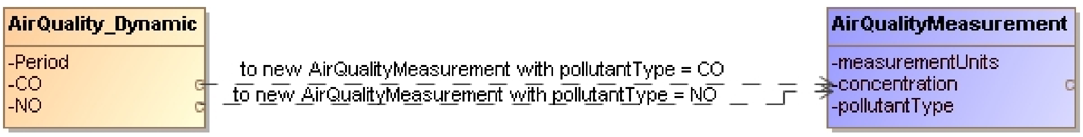
to AnotherClass.RelatedClass *with* anotherAttr = “value 2”;

from AClass.assocRole *where* anAttr == “value 11”

to YetAnotherClass.RelatedClass *with* yetAnotherAttr = “value 22”

A.1.7 “to” clause with “new”

Another case is where a single object instance maps to multiple core component object instances. For example one model groups pollutant measurements in a single instance while the core components have one instance per pollutant. A readable solution is to add a further keyword “new” into the “to” clause to illustrate that the presence of the attribute results in the creation of a separate core component object. This will typically be combined with a “with” clause as in the following example.



A similar approach may be used to map a single event to a pair of related events.

A.1.8 “where” clause with isKindOf()

It may be possible that attributes from an abstract base class could map to different core component elements depending on the concrete subclass. This may suggest that the core component model is not optimal, but if it does occur it may be represented by specifying “where isKindOf(concreteClassName)” in the attribute mappings for the abstract base class.

A.1.9 Further cases

- Mapping between numeric values and enumeration literals can be managed with the mechanisms described above, but the statements may be very long. For example consider the mapping between a numeric speed limit and an enumerated speed limit. The statement will take the form of:

  "where temporarySpeedLimit = 32, with speedLimit = LIMIT_20_MPH;

  where temporarySpeedLimit = 48, with speedLimit = LIMIT_30_MPH;" etc.

- The statements for enumeration to enumeration mappings are much more concise because there would be set of simple mappings from literal to literal.

- If a long statement would be too cluttered on the diagram, the dependency may simply say "see documentation for details", with the full statement being added to the documentation property of the dependency.

- The absence of an optional attribute may map to a specific value in core components (e.g. "false"). In this case use a "where" clause with the further keywords "present" or "absent" after the attribute name, i.e. "where [attribute] absent", or "where [attribute] present".

- A particularly difficult case is presented when submitted models take different approaches to state. For example, a model may include constructs to support *partial updates* of a concept. The core components do not support such a notion, and so although the individual attributes will have corresponding core components, the wrapper construct for the partial update will not. Supporting this kind of construct is out of scope for the core component mappings.

A.2 Use of the precise mappings

Derivation of precise mapping statements has the following benefits:

- They explain the mappings for human readers. In the example with Incident and Event, just mapping Incident to Event does not explain the thoughts behind this.

- The necessary rigour leads to improvements in the core component models. Although the core component derivation process is as objective as possible, it is easy to make a mistake unless the derivation process is rigorous.

- Translators can be implemented very directly from the mappings. The effort needed anyway for the translator will improve the core components and the understanding will be made available for future users.

The principal disadvantage is a high maintenance cost. If the core component mappings for a submitted model are registered, and either that model or the core components change, the mapping must be updated or removed. In the case of core component updates, every related mapping must be checked for changes.

Bibliography

[1] ISO/IEC 11179-3:2003, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

[2] ISO/IEC 11179-4:2004, *Information technology — Metadata registries (MDR) — Part 4: Formulation of data definitions*

[3] ISO/IEC 11179-5:2005, *Information technology — Metadata registries (MDR) — Part 5: Naming and identification principles*

[4] ISO/IEC 11404:2007, *Information technology — General-Purpose Datatypes (GPD)*

[5] ISO 14813-6:2009, *Intelligent transport systems — Reference model architectures for the ITS sector —Part 6: Data presentation using ASN.1*

[6] ISO 14817:2002, Transport information and control systems — Requirements for an ITS/TICS central Data Registry and ITS/TICS Data Dictionaries

[7] ISO/IEC 19501:2005, *Information technology — Open Distributed Processing — Unified Modeling Language (UML) Version 1.4.2*

[8] ISO/IEC TR 20943-1:2003, *Information technology — Procedures for achieving metadata registry content consistency — Part 1:Data elements*

[9] ISO/IEC TR 20943-3:2004, *Information technology — Procedures for achieving metadata registry content consistency — Part 3: Value domains*

[10] ISO/IEC PDTR 20943-5 [1] , *Information technology — Achieving Metadata Registry Content Consistency — Part 5: Semantic metadata mapping procedure*

[11] ISO 24531, *Intelligent transport systems — System architecture, taxonomy and terminology — Using XML in ITS standards, data registries and data dictionaries*

[12] Mott McDonald 221939/TD/01, ITS Data Registry Pilot Final Report, 19 September 2006

[13] TBG17 CCL (Core Component Library) Submission Guidelines and Procedures, 20 May 2008, Version 3.01

[14] FODOR, O., DELL'ERBA, M., RICCI, F. SPADA, A. and WERTHNER, H. *Conceptual Normalisation of XML Data for Interoperability in Tourism*, eCommerce and Tourism Research Laboratory ITC-irst, Italy

[15] Harmonise project, European Commission (EC), 5th Framework RTD Programme, contract number IST 2000-29329

[16] UN/CEFACT Core Components Technical Specification — Part 8 of the ebXML Framework, 15 November 2003, version 2.01

[1] To be published.

ICS 03.220.01; 35.240.60

Price based on 33 pages