

Reference number of working document: ISO/IEC JTC 1/SC 32 N **XXXX**

ISO/IEC JTC 1/SC 32 WG2 N **1625**

Date: 2012-01-03

Reference number of document: **ISO/IEC PDTR 20943-5**

Committee identification: ISO/IEC JTC 1/SC 32/WG 2

Secretariat: US

## **Information technology — Procedures for achieving metadata registry content consistency — Part 5: Semantic metadata mapping procedure**

### **Warning**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

### Copyright notice

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

*ISO copyright office*

*Case postale 56 • CH-1211 Geneva 20*

*Tel. + 41 22 749 01 11*

*Fax + 41 22 749 09 47*

*E-mail [copyright@iso.ch](mailto:copyright@iso.ch)*

*Web [www.iso.ch](http://www.iso.ch)*

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

# Contents

Page

Foreword .....	vii
Introduction.....	viii
1 Scope .....	1
1.1 Background.....	1
1.2 Purpose .....	1
1.3 Scope .....	1
2 Normative references .....	1
3 Definitions .....	2
3.1 Terms and definitions .....	2
3.2 List of abbreviations and acronyms.....	3
4 Semantic metadata mapping procedure.....	3
4.1 General .....	3
4.2 Data element mapping .....	4
4.2.1 Identifying metadata element sets (First process) .....	4
4.2.2 Grouping data elements (Second process).....	5
4.2.3 Semantic mapping (Third process) .....	6
4.3 Value mapping .....	7
4.3.1 Overview.....	7
4.3.2 Conversion .....	7
4.3.3 Recomposition.....	7
4.3.4 Examples .....	7
4.4 Types of heterogeneity .....	8
Annex A (Informative) Semantic metadata mapping and metadata registry .....	10
Annex B (Informative) Factoring.....	11
B.1 Introduction.....	11
B.2 Examples .....	11
B.3 Comparability.....	12
Bibliography.....	13

## Table of Figures

Figure 1 — Procedure for data element mapping .....	3
Figure 2 — Main and sub-processes for data element mapping .....	4
Figure 3 — Relationship between metadata mapping and registry .....	10

## Table of Tables

Table 1 — Analysing available metadata element sets .....	4
Table 2 — Example of grouping data elements by property.....	5
Table 3 — Semantic mapping of metadata .....	6
Table 4 — Examples of value mapping .....	7
Table 5 — Types of heterogeneity .....	9

## 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.

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/IEC TR 20943-5 was prepared by Technical Committee ISO/IEC JTC 1, Information technology, Subcommittee SC 32, Data management and interchange.

ISO/IEC TR 20943 consists of the following parts, under the general title Information technology — Procedures for achieving metadata registry content consistency:

- Part 1: Data elements
- Part 2: XML structured data
- Part 3: Value domains
- Part 4: Overview
- Part 5: Semantic metadata mapping procedure (this part)
- Part 6: Framework for generating ontologies

## Introduction

Usually, two systems do not share the same model, and that is because the categories represented in the models were not factored in the same way. This situation inhibits interoperability. For example, there may be two or more metadata element sets applicable to an information object. For example, metadata such as DC (Dublin Core), MARC (MACHINE Readable Cataloguing), and MODS (Metadata Object Description Schema) can be used to describe a book. Thus, a data element for an information object may be differently named due to the preferences of individual database developers. Consequently, data exchange among databases becomes difficult or almost impossible.

ISO/IEC 11179 provides a framework for achieving semantic interoperability of metadata between systems. A metadata registry based on ISO/IEC 11179 offers a good way to secure interoperability among databases. However, there are many metadata sets which are not following ISO/IEC 11179. In order to mediate among plural data elements already developed or used, other measures are necessary. In general, interoperability may be achieved through conformity to some set of provisions. For example, metadata crosswalk is the most commonly used way to map a data element to another data element. However, the metadata crosswalk has poor semantics; it provides a simple one-to-one mapping table among data elements without any explanation about the semantic relationship. Therefore, the metadata crosswalk needs to be elaborated in order to give semantics and to cover cases other than one-to-one mapping. The basic concept of ISO/IEC 11179 can still be applicable to the improvement of semantic metadata crosswalk because it addresses the semantics of metadata and naming principles for data elements.

Some other approaches have been tried to provide guidelines or a model for harmonization of metadata and data in especially transport industry. A Model for Semantic Equivalence Discovery for Harmonizing Master Data, presented at OTM 2009 Workshops, suggests a model for semantic mapping of master data. ISO TR 25100 Intelligent transport systems — Systems architecture — Harmonization of ITS data concept reviews four harmonization processes applicable to an ITS/TICS central data registry and ITS/TICS data dictionaries and recommends a practical guidelines. Both literatures are examples restricted within transport industry but including rationale for the need of semantic metadata mapping in general.

This Technical Report describes a semantic metadata mapping procedure (SMMP), which can maximize the interoperability among data elements. The procedure consists of three main processes: identifying metadata element sets, grouping data elements, and semantic mapping. This Technical Report also includes types of value mapping.

# Information technology — Procedures for achieving metadata registry content consistency —

## Part 5: Semantic metadata mapping procedure

### 1 Scope

#### 1.1 Background

Metadata differs in many ways, yet for the purposes of data and semantic interoperability, those differences must be bridged or harmonized. For example, especially data elements having different names even though they have the same meanings may cause a data discrepancy problem when such data are shared or interchanged. Thus, semantic metadata mapping is required to mediate among these data elements to allow sharing or interoperable use. A metadata crosswalk is the most commonly used method to map a data element to another data element. However, it has poor semantics because it is meaningful only for simple one-to-one mapping. Therefore, the metadata crosswalk needs to be elaborated in order to have semantics and to cover cases other than one-to-one mapping.

#### 1.2 Purpose

The purpose of this Technical Report is to set up a procedure for making metadata crosswalks that conform to the ISO/IEC 11179 standard, subsequently improving the semantic harmonization of metadata: DE and VD.

#### 1.3 Scope

This Technical Report describes a semantic metadata mapping procedure (SMMP), which can maximize the interoperability among data elements. The procedure of data element mapping consists of three main processes, which are divided into nine sub-processes. The main processes are identifying metadata element sets, grouping data elements, and semantic mapping. This Technical Report includes a simple example to explain each process.

This Technical Report is recommended for use in a specific subject domain because the procedure can be more meaningful when a specific information object is concerned.

This Technical Report addresses interoperability of values as well.

### 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 704:2009, *Terminology work — Principles and methods*

ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO/IEC 11179-1:2004, *Information technology — Metadata registries (MDR) — Part 1: Framework for the specification and standardization of data elements*

ISO/IEC 11179-3:2012, *Information technology — Metadata registries (MDR) — Part 3: Registry Metamodel and basic attributes*

ISO/IEC 11179-5:2005, *Information technology — Metadata registries (MDR) — Part 5: Naming and identification principles for data elements*

## **3 Definitions**

### **3.1 Terms and definitions**

For the purposes of this document, the terms and definitions given elsewhere in ISO/IEC 11179 and the following apply.

#### **3.1.1**

##### **crosswalk**

mapping of the elements, semantics, and syntax from one metadata scheme to those of another

[NISO, 2004]

#### **3.1.2**

##### **complicated difference**

type of heterogeneity that cannot be harmonized

#### **3.1.3**

##### **domain difference**

type of heterogeneity arising from the different kinds of contexts or cultures

#### **3.1.4**

##### **hierarchical difference**

type of heterogeneity arising from the different levels of details

#### **3.1.5**

##### **interoperability**

ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality

[ISO/IEC 2382-1:1993]

#### **3.1.6**

##### **lexical difference**

type of heterogeneity arising from different appearances

#### **3.1.7**

##### **metadata element set**

small and fundamental group of data elements through which resources can be described and catalogued in a domain

#### **3.1.8**

##### **primary data element**

data element to which other data elements are mapped

#### **3.1.9**

##### **syntactic difference**

type of heterogeneity arising from varying arrangement of parts

## 3.2 List of abbreviations and acronyms

### 3.2.1

#### **DC**

Dublin Core

### 3.2.2

#### **MARC**

MAchine Readable Cataloguing

### 3.2.3

#### **MDR**

MetaData Registry

### 3.2.4

#### **MODS**

Metadata Object Description Schema

### 3.2.5

#### **OEBPS**

Open EBook Publication Structure

### 3.2.6

#### **SMMP**

Semantic Metadata Mapping Procedure

### 3.2.7

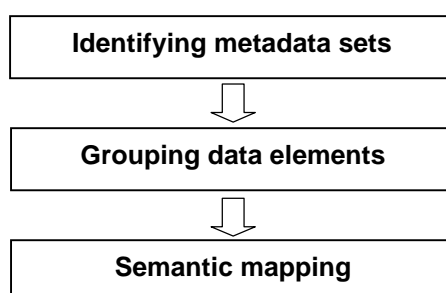
#### **TEI**

Text Encoding Initiative

## 4 Semantic metadata mapping procedure

### 4.1 General

In this Technical Report, metadata mapping covers data element mapping and value mapping. The procedure for data element mapping consists of three main processes as shown in Figure 1.



**Figure 1 — Procedure for data element mapping**

The first process is to identify metadata element sets required to be mapped. It is necessary to survey available metadata element sets (in a specific domain).

The second process is to group data elements obtained from the identified metadata element sets, including four consecutive sub-processes namely, finding objects, grouping all data elements by object, finding their properties, and grouping all data elements by property.

The last process involves mapping data elements semantically. In this process, it is necessary to arrange all data elements into a table. Notes on the accuracy of matching are included in every slot of the table. A recommended set of metadata can also be provided in the process for guiding future standardization.

Figure 2 shows all sub-processes related to corresponding main processes.

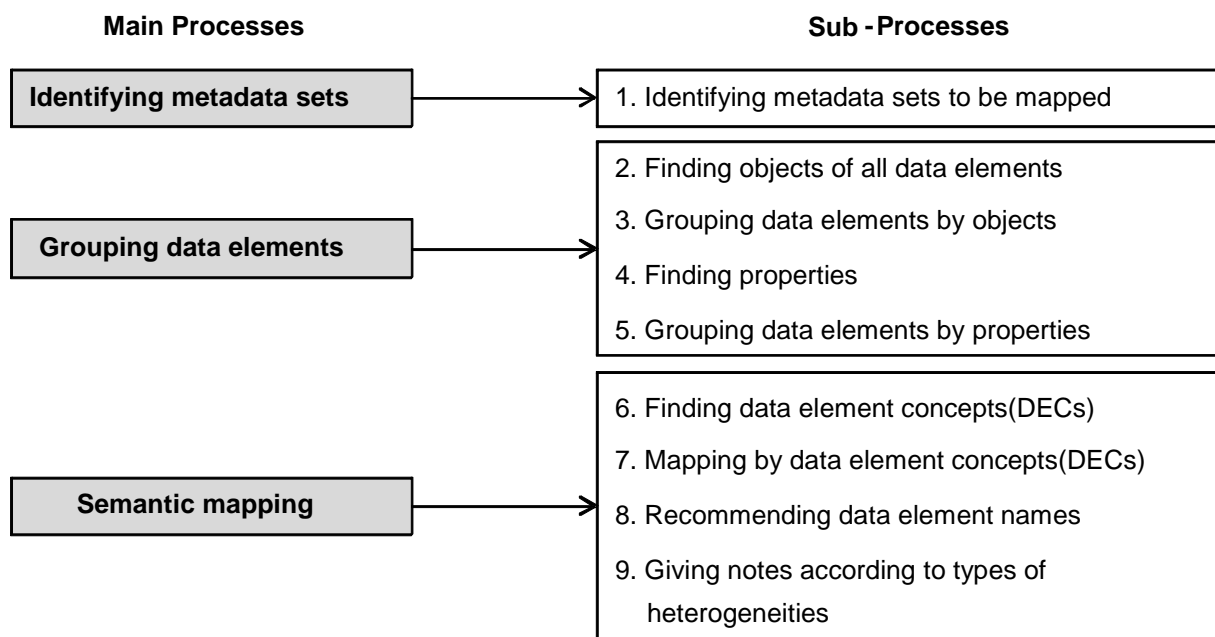


Figure 2 — Main and sub-processes for data element mapping

## 4.2 Data element mapping

### 4.2.1 Identifying metadata element sets (First process)

#### 4.2.1.1 Method

First, it is necessary to collect available metadata element sets and to identify candidate metadata element sets to be mapped. Then, what the domain or service database is should be checked, how many numbers of fields should be counted, and whether sample data exists or not should be checked. Who or which organization has the authority over each metadata element set should also be checked.

If the metadata element set or target object is not suitable for mapping, it may not be chosen.

#### 4.2.1.2 Examples

- Domain: e-book cataloging
- Available metadata element sets: OEBPS, MODS and TEI

Table 1 — Analysing available metadata element sets

Metadata element set name	OEBPS	MODS	TEI header
Domain or service database	Description of electronic book	Description of library resources	Encoding methods for machine-readable texts
Number of fields	15	About 60	Over 20
Sample data	yes	No	Yes
Authority	Open eBook Forum	Library of Congress	TEI Consortium

## 4.2.2 Grouping data elements (Second process)

### 4.2.2.1 Method

The next process is to group data elements by object class, and then to find properties involved in the objects and sub-grouped data elements by properties.

For convenience, it is helpful to select a primary data element among the collected data elements and aggregate data elements by the primary data element. The simplest or the highest-level data element is recommended to be the primary data element.

All data elements included in the candidate metadata element sets should be aggregated by property. Data elements relatively less important may be eliminated. Some data elements, which cannot be grouped, are supposed to be set aside.

In this process, metadata experts should perform the work along with domain experts.

### 4.2.2.2 Examples

In the sample object class, the e-Book has plural properties as shown below.

- Object class: e-Book
- Properties: title, author, subject, ..., edition

**Table 2 — Example of grouping data elements by property**

OEBPS*	MODS	TEI
Title	Title subTitle <i>partNumber</i> <i>partName</i> <i>nonSort</i>	title seriesStmt:title <i>seriesStmt:idno</i>
Creator(role) Creator(file-as)	name:role name:namePart <i>name:displayForm</i> <i>name:affiliation</i> <i>name:discription</i>	author
Subject	Topic classification <i>catographics</i> <i>occupation</i>	keyword classCode <i>catRef</i>
...	...	...
(no data element)	Edition	<i>fileDesc_editionStmt_date</i> <i>fileDesc_editionStmt_edition</i> <i>fileDesc_editionStmt_respStmt</i> <i>fileDesc_editionStmt_respStmt_name</i> <i>fileDesc_editionStmt_respStmt_resp</i>

\* primary data elements

Similar properties of MODS and TEI are grouped according to those of the primary data elements from OEBPS. In the table, the italicized parts refer to properties considered less important in the target application domain.

### 4.2.3 Semantic mapping (Third process)

#### 4.2.3.1 Method

After identifying object classes and properties hidden in and related to all the primary data elements, we can create common Data Element Concepts (DECs) according to ISO/IEC 11179-1 and 3.

The third process starts from finding common data element concepts in each group of data elements based on objects and properties found in the second process. The existence of a domain taxonomy or ontology will make it easier to construct the common DECs.

Finally, all candidate data elements are arranged into a table by the common DECs. Types of heterogeneity can be described in the table. The types consist of six categories. (See detail in 4.4)

- Same, no difference: no description
- Hierarchical difference: H/gen(Generalization), H/spe(Specialization), H/com(Composition), H/dec(Decomposition)
- Domain difference: D
- Lexical difference: L/syn(Synonyms), L/abb(Abbreviation), L/acc(Acronyms), L/cas(Case sensitivity), L/lan(Language), L/var(Variation)
- Syntactic difference: S/ord(Ordering), S/del(Delimiters), S/mis(Missing)
- Complicated difference: C

A recommended data element can be provided for the future standardization work.

#### 4.2.3.2 Examples

The data element concepts found can be shown as follows:

- DECs: ebookTitle, ebookAuthor, ebookSubject, ebookEdition

Finally, we can create DECs according to ISO/IEC 11179-1. New DECs should be also created for data elements that could not be grouped during the second process.

Table 3 shows the final result obtained through the procedure. The common DECs are described in the first column while the recommended data elements are done in the right end column. Between them are data elements from the candidate metadata element sets.

**Table 3 — Semantic mapping of metadata**

Common DEC	OEBPS		MODS		TEI		Recommended DE*
<b>ebookTitle</b>	Title		Title	H/dec	Title	H/dec	<b>ebookTitle</b>
			Subtitle	H/dec	seriesStmt:title	H/dec S/del	
<b>ebookAuthor</b>	Creator(role)	C	name:role	C	Author	S/mis	<b>ebookAuthorName</b>
	Creator(file-as)	D	name:namePart	D			
<b>ebookSubject</b>	Subject		Topic	L/syn	Keyword	L/syn	<b>ebookSubject</b>
			Classification	L/syn	Class	L/syn	
<b>ebookEdition</b>			Edition	L/cas S/mis	Edition	L/cas S/mis	<b>ebookEditionNumber</b>

\* Recommended DE is provided according to the DE naming rule from ISO/IEC 11179-5.

## 4.3 Value mapping

### 4.3.1 Overview

Even though the meanings of data elements are mapped, and even if the data type of the value domains is the same, all of the value domains of data elements need to be checked and harmonized in order to interchange real data. When the data types are not the same, the representation of a value may be converted from one format to the other.

The conversion of the value from one Value Domain to another is applicable only when the Conceptual Domain of each Value Domain to be mapped is same. If the Conceptual Domains are different, value mapping may only be performed partially or may not be possible. The values should be mapped only if the Value Meaning Identifiers are the same.

### 4.3.2 Conversion

Conversion can only apply to determinate values.

- **Code conversion** - In case of using different code sets sharing same conceptual domain for the value domain, a code set is converted to the other code set.
- **Letter conversion** - In case of using different characters such as an abbreviation, acronym, different case, lexical variation, or different delimiters for the Value Domain Value, the Value is converted to a preferred representation.
- **Unit conversion** - In case of using numeric value for the value domain, the numeric value is converted using an appropriate conversion rule.
- **Generalization** - If two or more values in one value domain can be represented by a broader value in the other value domain, these values are converted to the broader value.

### 4.3.3 Recomposition

- **Composition** - If a value is separated into two or more data elements, those values are composed into one value of an appropriate data element.
- **Decomposition** - If a value should be separated into two or more data elements, the value is decomposed into two or more values of appropriate data elements.
- **Order change** - If a value is composed of two or more words which are ordered differently, the order of the words are changed to have the same order.

### 4.3.4 Examples

Table 4 shows the examples of value mapping.

**Table 4 — Examples of value mapping**

Mapping types		Examples		
Conversion	Code conversion	KOR to KR		
	Letter conversion	KATS to Korea Agency for Technology and Standards		
	Unit conversion	1 in to 2.54 cm		
	Generalization	(Married, Widowed and Divorced) to Married		
Recomposition	Composition	First Name: John Given Name: Kennedy	to	Full Name: John Kennedy
	Decomposition	Full Name: John Kennedy	to	First Name: John Given Name: Kennedy
	Order change	John Kennedy to Kennedy, John		

## 4.4 Types of heterogeneity

Type of heterogeneity can be derived from the factoring (see Annex B for more information).

Factoring for the description of a datum can be applied to data elements, where the descriptions are on sets of semantically and representationally related data rather than on a single datum. It is clear that the attributes from ISO/IEC 11179 represent a factorization of the description of data elements (not a particular data element, but data elements in general). The categories are based on terminological, computational, and representational considerations. To make this work, attributes may be structured linearly, hierarchically, or in some other grouping. The choice is one that allows the user to make comparisons easily.

The steps in factoring a description are:

- Step #1: Factor the data description according to terminological principles. Express the description as a reification of identified attributes. For the data element described as “The daily temperature, in Celsius, measured at 10:00 on Roosevelt Island, NY, USA, measured by thermometer XYZ, to 0.1 degrees precision, with accuracy 0.2”, the attributes shown below are identified.
- Step #2: Express the attributes.
- Step #3: Determine which attributes are used for comparison with data from another source. Harmonization is dependent upon these characteristics.

There are four possibilities to what happens when the factor categories are the same but the attributes don't have the same data. They are:

- 1) 2 concepts are generalized to a concept set that subsumes both (generalization)
- 2) 2 concepts are specialized to a concept set that both subsume (specialization)
- 3) A new concept is produced from 2 separate ones that both are related to
- 4) The 2 concepts are incommensurable (complicated)

Based on the above, the types of semantic heterogeneity of metadata can be classified into six categories: no difference, hierarchical difference, domain difference, lexical difference, syntactic difference, and complicated difference.

Identical data elements can be mapped using one-to-one mapping.

Hierarchical Difference is caused by different levels of details such as generalization/specialization and composition/decomposition. For example, a general term 'price' can be specified to 'retail price', or 'whole sale price'; and a person's 'name' can be composed of 'first name', 'middle name', and 'family name'. Data elements having hierarchical differences should be mapped using the higher level or composed terms if required. (See more details in ISO/IEC 20943-1:2003)

Domain Difference is caused by different contexts or cultures, e.g. summary vs. synopsis. In this case, one-to-one mapping is possible.

Lexical Difference refers to the different appearance of a data element related to synonyms, abbreviations, acronyms, case sensitivity, languages, or variation. This kind of heterogeneity is solved by one-to-one mapping.

Syntactic Difference arises from the varying arrangement of parts within a data element name. The order of words, the type of delimiter, and missing words can cause heterogeneity between semantically same data elements, e.g. ordering (family name : name (family)), delimiters (Family-Name : FamilyName), and missing (classification code : classification). This type of heterogeneity can likewise be solved by one-to-one mapping.

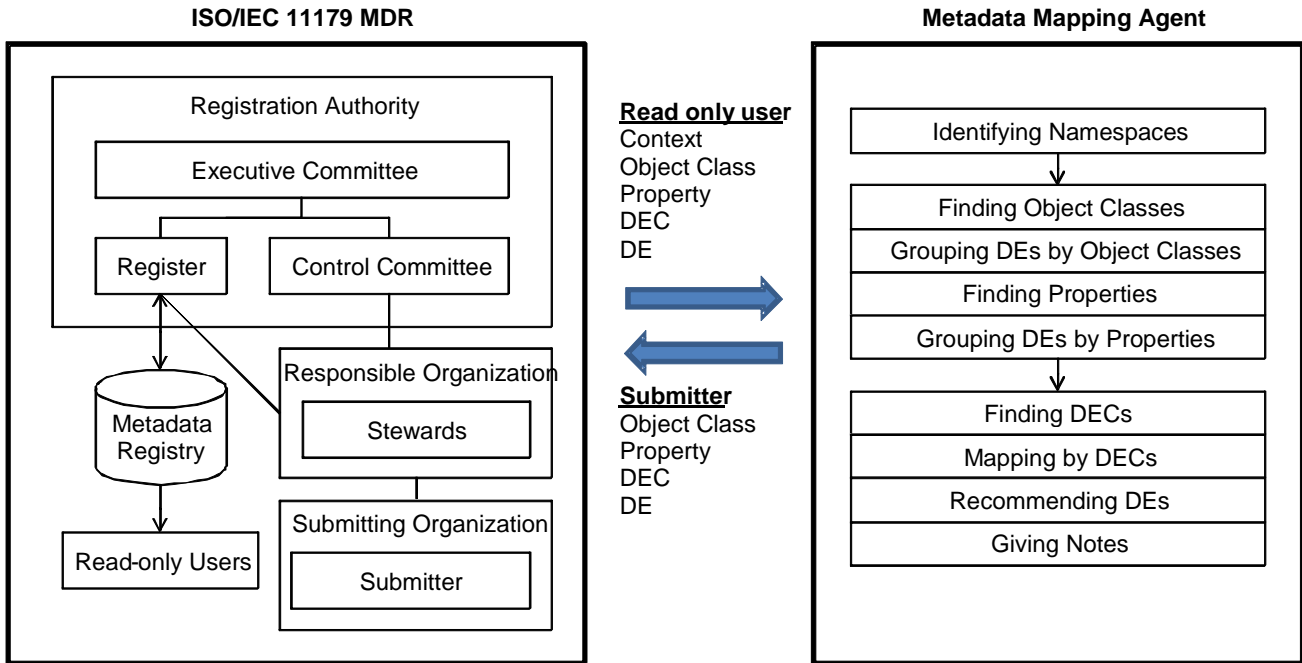
Complicated difference is the type of heterogeneity that cannot be solved without human intervention.

Table 5 — Types of heterogeneity

Type	Sub-Type	Mark	Examples
	Ways of harmonization (Types of mapping)		
Same	Same		
	One-to-one mapping		
Hierarchical	Generalization Specialization	H/gen H/spe	Price Retail price, Wholesale price, ...
	One-to-one mapping (dumb down)		
	Composition Decomposition	H/com H/dec	Name Family name, given name, ...
	One-to-many or many-to-one mapping (if required)		
Domain	Domain	D	Summary : Synopsis
	One-to-one mapping (if required)		
Lexical	Synonyms	L/syn	First name : Given name
	Abbreviation	L/abb	Address : Addr.
	Acronyms	L/acr	Serial Number :SN
	Case sensitivity	L/cas	Address : ADDRESS
	Language	L/lan	Name : 이름
	Variation	L/var	Color : Colour
	One-to-one mapping		
Syntactic	Ordering	S/ord	Family name : Name (family)
	Delimiters	S/del	Family-name : Family_name
	Missing	S/mis	Author name : Author
	One-to-one mapping		
Complicated	Complicated	C	
	Mapping is impossible.		

## Annex A (Informative) Semantic metadata mapping and metadata registry

When semantic metadata mapping is performed among metadata registries based on ISO/IEC 11179, the mapping agent can act as a read-only user and a submitter of the metadata registries. During the first and second processes, it uses contexts, object classes, properties, data element concepts, and data elements out of the metadata registries. It can also submit new object classes, properties, data element concepts, and data elements newly obtained in the last process.



**Figure 3 — Relationship between metadata mapping and registry**

## Annex B (Informative) Factoring

### B.1 Introduction

Factoring is the process of assigning the concepts in some description into pre-defined classes. Sometimes more than one concept is placed in a single class, and sometimes there is a one to one correspondence; it depends on the complexity of the description to be factored and the available classes. This technical report provides the classes into which data element descriptions are sub-divided; the technical report does not address how one factors a description into this sub-division is to take place.

Given a complex description of some item, there are no robust rules for deciding which part of a description goes into a particular class. For descriptions of data elements, this problem is particularly acute for those implementing a metadata registry based on ISO/IEC 11179. Some of the stated purposes of this technical report are to promote comparability, interoperability, and the ability to harmonize data across systems. Yet, even though all implementers of ISO/IEC 11179 use the same classes to store data element descriptions, factoring those descriptions into these classes is still a source of variability.

We will look at this problem from the point of view of an example. Some remedies for those implementing ISO/IEC 11179 conforming metadata registries will become apparent.

### B.2 Examples

For the purposes of the analysis of the example to follow, we will use the following classes for factoring a description of a data element from ISO/IEC 11179-3. These are

- Object class
- Property
- Conceptual domain
- Value domain
- Datatype

There is no loss of generality by excluding other classes, since these are the main classes for describing data elements. In particular, the data element concept is a combination of the object class and the property.

With respect to ISO/IEC 11179, each of the classes listed above are considered to be independent ideas. That is, there is no overlap between them as far as their descriptive purposes are concerned. The object class describes the objects for which data are (or to be) collected (Example: people). The property is that feature of those objects that is to be determined (Example: sex). The conceptual domain provides the set of determinants, the possible ways a determination of the property can be resolved for each object (Example: {female, male}). The value domain provides the way each determinant will be represented (Example: f for female, and m for male). The datatype describes the computations that are allowable for the determinants in the conceptual domain (Example: the 'state' datatype in ISO/IEC 11404).

This set of classes seems self-evident, and it is one of the reasons that ISO/IEC 11179 was developed in the way that it is. On the other hand, examples show that it is not straightforward as to which sections of a description go in particular classes. Take the following description of a data element: "Takeoff time of the president's aircraft in ISO 8601 format". We want to determine what parts of this description go in which classes. It turns out there are several ways this can be done, and none of them are any better than the others. In fact, the choice as to which factoring is the best is determined by the analyses intended.

Here are some possible factorings of the given data element description:

**Factoring #1**

Object class: takeoffs of president's aircraft  
Property: time  
Conceptual domain: date/times  
Value domain: ISO 860 format  
Datatype: date-time (see ISO/IEC 11404)

**Factoring #2**

Object class: president's aircraft  
Property: takeoff time  
Conceptual domain: date/times  
Value domain: ISO 860 format  
Datatype: date-time (see ISO/IEC 11404)

**Factoring #3**

Object class: aircraft  
Property: president's takeoff time  
Conceptual domain: date/times  
Value domain: ISO 860 format  
Datatype: date-time (see ISO/IEC 11404)

These examples show that in general there is no canonical way to factor a description. Often, there are many possibilities. The particular way problems arose here was between the object class and property, however there are other possibilities as well.

Examination of each of these possible factorizations shows that the unit of analysis changes in each case. This is reflected in the changes to the object class. As the overall description remains the same, the property then must change. Notice, too, that the data element concept does not change at all among these cases, and therefore the further factoring into object class and property provides more semantics.

On the other hand, it is possible to see the changes from the point of view of the property instead. This means what is being measured or determined is changing. Since the description doesn't change, the object class must change as a result.

### **B.3 Comparability**

The variability in how a description might be factored has a huge impact on comparability. If the object classes between two data elements are not the same, the chance for them to be comparable, let alone harmonized, goes down. The same is true for differences between properties. Luckily, the various ways to factor a description also point to the ways in which the semantics of a data element can be represented. If the managers of the metadata registry recognize the power of interpreting the semantics of a data element in more than one way, the metadata registry can become a very powerful mechanism.

Taking the example we looked at above, if each of the three possible factorizations is stored as a separate (but related, because each describes the same data) data element description, this vastly increases the possibility for achieving or finding comparable data elements in other systems. It may seem counterintuitive that the same description of a data element could provide different semantics, but the needs of data analysis may be different than the desires of data design or production. Understanding the possibility of analyzing data in ways other than they were originally designed provides a rationale for factoring some data element descriptions multiple ways.

For the managers of the metadata registry, this will mean more work, but the efforts needed for data comparability and harmonization are not easy. The view presented here provides a new way to view that process. It takes advantage of the flexible structure in the ISO/IEC 11179 standard, and it aims to maximize the use of registries based on that standard.

## Bibliography

- [1] DC (Doblin Core)  
<http://dublincore.org/documents/dces/>
- [2] MARC (MAchine Readable Cataloguing)  
<http://www.loc.gov/marc/>
- [3] MODS (Metadata Object Description Schema)  
<http://www.loc.gov/standards/mods/>
- [4] OEBPS (Open eBook Publication Structure)  
<http://old.idpf.org/oebps/oebps1.2/index.htm>
- [5] TEI (Text Encoding Initiative)  
<http://www.tei-c.org/index.xml>
- [6] A Model for Semantic Equivalence Discovery for Harmonizing Master Data, Baba Piprani, OTM 2009 Workshops, LNCS 5872, pp. 649–658, 2009
- [7] ISO TR 25100:2010 Intelligent transport systems — Systems architecture — Harmonization of ITS data concept