TECHNICAL REPORT

ISO/IEC TR 20943-5

First edition
2013-11-01

Information technology — Procedures for achieving metadata registry content consistency —

Part 5: Metadata mapping procedure

Technologies de l'information — Procédures pour réaliser la consistance du contenu de l'enregistrement des métadonnées —

Partie 5: Procédure de mappage des métadonnées

Reference number
ISO/IEC TR 20943-5:2013(E)

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

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, when the joint 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 to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 20943-5 was prepared by Joint 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 3: Value domains*

— *Part 5: Metadata mapping procedure*

— *Part 6: Framework for generating ontologies*

Introduction

Metadata differs in many ways, yet for the purposes of data and semantic interoperability, those differences must be bridged or harmonized. For example, synonyms may cause discrepancies. If two data elements mean the same thing but have different names, that difference may cause data under each to be flagged as incompatible when they aren't.

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. There may be two or more metadata element sets applicable to an information object. For example, metadata schemas, such as DC (Dublin Core),[10] MARC (MAchine Readable Cataloguing),[11] and MODS (Metadata Object Description Schema)[12] can be used to describe books. So, the result might be significantly different data elements in different schemas, even though the schemas describe the same objects.

ISO/IEC 11179 provides a framework for achieving 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 is meaningful only for exact matching; it provides a simple one-to-one mapping table between data elements without any explanation about the relationship. Therefore, the metadata crosswalk needs to be elaborated in order to cover cases other than exact matching.

Some other approaches have been tried to provide guidelines or a model for harmonization of metadata and data in especially the transport industry. A Model for Semantic Equivalence Discovery for Harmonizing Master Data,[9] presented at OTM 2009 Workshops, suggests a model for semantic mapping of master data. ISO/TR 25100[8] reviews four harmonization processes applicable to an ITS/TICS central data registry and ITS/TICS data dictionaries and recommends practical guidelines. Both literatures are examples restricted within the transport industry but including a rationale for the need of metadata mapping in general.

This part of ISO/IEC TR 20943 describes a metadata mapping procedure (MMP), which can maximize the interoperability among metadata.

Information technology — Procedures for achieving metadata registry content consistency —

Part 5:
Metadata mapping procedure

1 Scope

The purpose of this part of ISO/IEC TR 20943 is to describe a procedure for establishing metadata crosswalks based on the ISO/IEC 11179 series, subsequently improving mapping quality between metadata.

Therefore, this part of ISO/IEC TR 20943 describes a metadata mapping procedure (MMP), which can maximize the interoperability among ISO/IEC 11179-based registries through achieving metadata registry content consistency.

2 Normative references

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

ISO/IEC 11179-1:2004, *Information technology — Metadata registries (MDR) — Part 1: Framework*

ISO/IEC 11179-3:2013, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

3 Terms and definitions

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

3.1
crosswalk
mapping (3.8) of the elements, semantics, and syntax from one metadata scheme to those of another

Note 1 to entry: Definition taken from *Understanding Metadata*, 2004, NISO Press.[15]

3.2
complicated difference
type of *heterogeneity* (3.4) that cannot be harmonized

3.3
domain difference
type of *heterogeneity* (3.4) arising from different kinds of contexts or cultures

3.4
heterogeneity
difference arising from different descriptions of the same concept

3.5
hierarchical difference
type of *heterogeneity* (3.4) arising from the different levels of details

3.6
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

Note 1 to entry: Definition taken from *Understanding Metadata*, 2004, NISO Press.[15]

3.7
lexical difference
type of *heterogeneity* (3.4) arising from different designations

[SOURCE: ISO 1087-1:2000, 3.4.1]

3.8
mapping
correspondence between instances of one model and instances of another model that represent the same meaning

[SOURCE: ISO/TS 18876-1:2003, 3.1.14]

3.9
metadata element set
small and fundamental group of data elements or data element concepts through which resources can be described and catalogued in a domain

EXAMPLE DC (Dublin Core), MARC (MAchine Readable Cataloguing), MODS (Metadata Object Description Schema), etc.

3.10
primary metadata element set
metadata element set (3.9) to which other metadata element sets are grouped

3.11
secondary metadata element set
metadata element set (3.9) which are grouped according to the *primary metadata element set* (3.10)

3.12
syntactic difference
type of *heterogeneity* (3.4) arising from varying arrangement of parts

4 Abbreviated terms

DC Dublin Core

DCMI Dublin Core Metadata Initiative

DE Data Element

DEC Data Element Concept

MARC MAchine Readable Cataloguing

MDR MetaData Registry

MMP Metadata Mapping Procedure

MODS Metadata Object Description Schema

NISO National Information Standards Organization (NISO is a USA organization: www.niso.org)

OEBPS Open EBook Publication Structure

TEI Text Encoding Initiative

5 Metadata mapping procedure

5.1 Overview

This part of ISO/IEC TR 20943 covers data element concept mapping and conceptual domain mapping. The mapping procedure for data element concept, described in this part of ISO/IEC TR 20943, is applicable to data element mapping, while the mapping procedure for conceptual domain includes value domain mapping.

5.2 Data element concept mapping

5.2.1 Overview

The procedure for data element concept mapping consists of three main processes as shown in Figure 1.

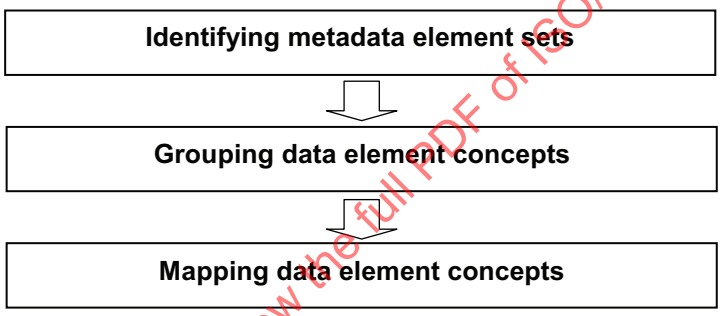


Figure 1 — Procedure for data element concept mapping

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 element concepts obtained from the identified metadata element sets. This includes four consecutive sub-processes: finding object classes, grouping all data element concepts by object class, finding their properties, and grouping all data element concepts by property.

The last process involves mapping data element concepts. In this process, it is necessary to arrange all data element concepts into a table. Notes on the type of mapping are included in every slot of the table.

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

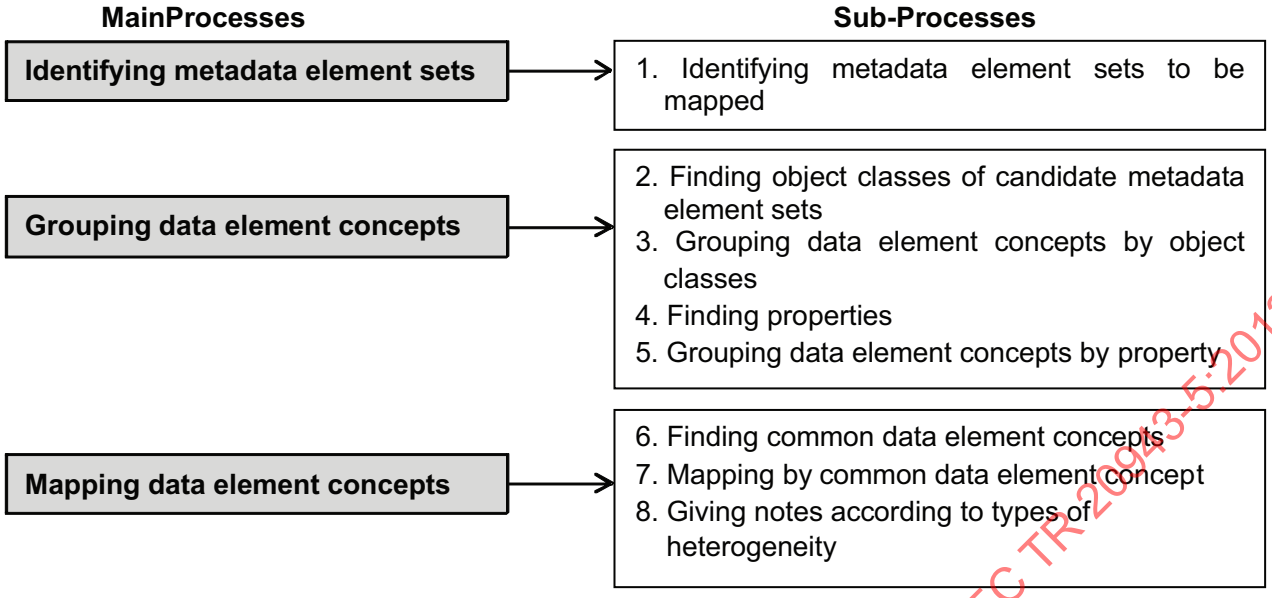


Figure 2 — Main and sub-processes for data element concept mapping

5.2.2 Identifying metadata element sets (First process)

5.2.2.1 Method

First, it is necessary to collect available metadata element sets and to identify candidate metadata element sets to be mapped. Then, what the purpose of the metadata element set 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.

5.2.2.2 Examples

— Domain: e-book cataloguing

— Available metadata element sets: OEBPS,[13] DC,[10] MODS[12] and TEI[14]

Table 1 — Analysing available metadata element sets

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

5.2.3 Grouping data element concepts (Second process)

5.2.3.1 Method

The next process is to group data element concepts by object class, and then to find properties associated with the objects and sub-group the data element concepts by property.

For convenience, it is helpful to select a primary metadata element set among the candidate metadata element sets and aggregate them by the primary metadata element set. The metadata element set having the least number of elements is recommended to be the primary metadata element set.

All data element concepts included in the candidate metadata element sets should be aggregated by property. 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.

5.2.3.2 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 element concepts by property

Primary meta-data element set	Secondary metadata element set		
DC	**OEBPS**	**MODS**	**TEI**
Title	Title	Title subTitle partNumber partName nonSort	title seriesStmt:title seriesStmt:idno
Creator	Creator(role) Creator(file-as)	name:role name:namePart name:displayForm name:affiliation name:description	author
Subject	Subject	Topic classification catographics occupation	keyword classCode catRef
…	…	…	…
(none)	(none)	Edition	fileDesc_editionStmt_date fileDesc_editionStmt_edition fileDesc_editionStmt_respStmt fileDesc_editionStmt_respStmt_name fileDesc_editionStmt_respStmt_resp

Similar properties of OEBPS, MODS and TEI are grouped according to those of the primary metadata element set.

5.2.4 Mapping data element concept (Third process)

5.2.4.1 Method

After identifying object classes and properties related to the primary metadata element set, common Data Element Concepts (DECs) can be created according to ISO/IEC 11179-1 and ISO/IEC 11179-3.

The third process starts from finding common data element concepts in each group of data element concepts based on object classes 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 element concepts are arranged into a table by the common DECs. Types of heterogeneity can be described in the table. The types consist of six categories as Table 3. Derivation of types of heterogeneity is explained in Annex B.

Table 3 — Types of heterogeneity

Type number	Type	Sub-Type number	Sub-Type	Mark	Examples
		Ways of harmonization (Types of mapping)			
1	Identical	1.1	Identical		
		One-to-one mapping			
2	Hierarchical	2.1	Generalization	H/gen	Retail price, Wholesale price→Price
		2.2	Specialization	H/spe	Price→Retail price, Wholesale
		One-to-one mapping (dumb down)			
		2.3	Composition	H/com	Family name, given name→Name
		2.4	Decomposition	H/dec	Name→Family name, given name
		One-to-many or many-to-one mapping (if required)			
3	Domain	3.1	Domain	D	Summary: Synopsis
		One-to-one mapping (if required)			
4	Lexical	4.1	Synonyms	L/syn	First name: Given name
		4.2	Abbreviation	L/abb	Address:Addr.
		4.3	Acronyms	L/acr	Serial Number:SN
		4.4	Case sensitivity	L/cas	Address: ADDRESS
		4.5	Language	L/lan	Name: 이름
		4.6	Variation	L/var	Color: Colour
		One-to-one mapping			
5	Syntactic	5.1	Ordering	S/ord	Family name: Name (family)
		5.2	Delimiters	S/del	Family-name: Family_name
		5.3	Missing	S/mis	Author name: Author
		One-to-one mapping			
6	Complicated	6.1	Complicated	C	
		Mapping is impossible.			

5.2.4.2 Examples

The data element concepts found can be shown as follows:

— DECs: ebookTitle, ebookAuthor, ebookSubject

Finally, we can create DECs according to ISO/IEC 11179-1. A new DEC, ebookEdition, is also created for Edition that could not be grouped during the second process.

Table 4 shows the final result obtained through the procedure. The first column is common DECs, the follows are data element concepts from the candidate metadata element sets.

Table 4 — Examples of data element concept mapping

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

5.3 Conceptual domain mapping

5.3.1 Overview

For data element concepts which have beenmapped, conceptual domain and value domainmustbe checked and harmonized as well. In this case mapping type comprises conversion and recomposition.

5.3.2 Conversion

— 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, theyare 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 conceptual domain (or value domain) can be represented by a broader value in the other conceptual domain (or value domain), these values are converted to the broader value.

5.3.3 Recomposition

— Composition - If a value is separated into two or more data element concepts, those values are composed into one value of an appropriate data element concept.

— 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.

5.3.4 Examples

Table 5 shows the examples of conceptual domain mapping.

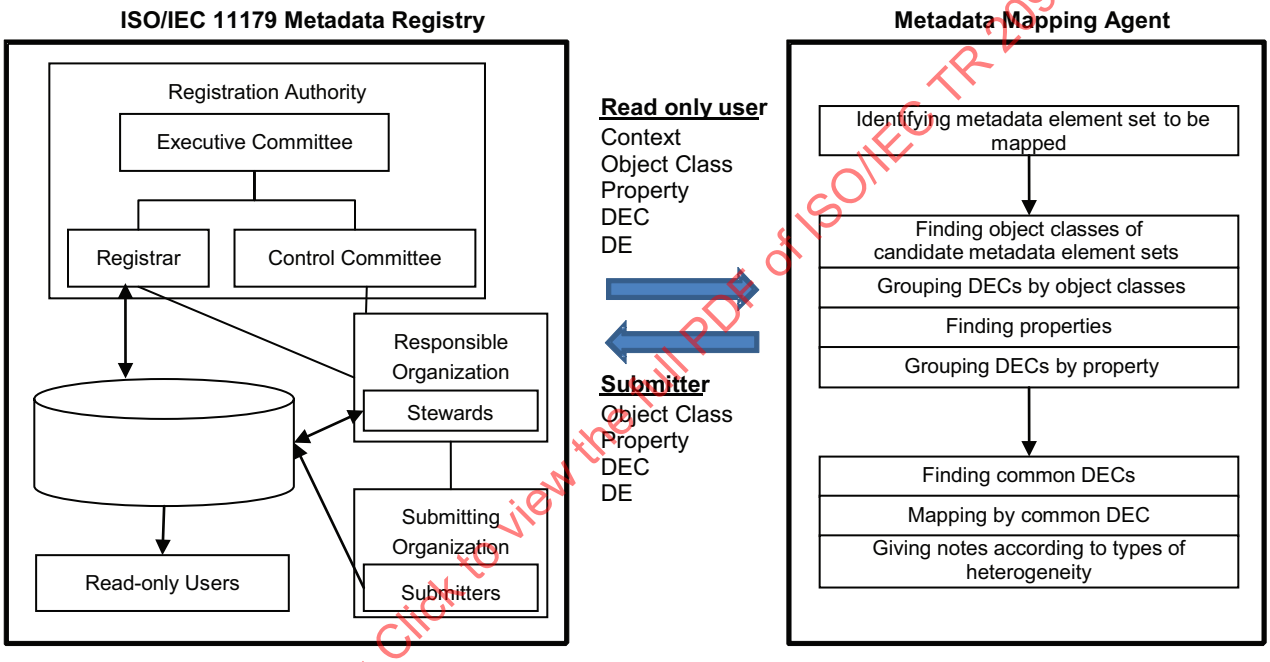
Table 5 — Examples of conceptual domain mapping

Mapping types		**Examples**		
Conversion	Code conversion	KOR to KR		
	Letter conversion	IJME to Int. J MechEng		
	Unit conversion	1 in to 2,54 cm		
	Generalization	(father, mother) to parent		
Recomposition	Composition	First Name: John Family Name: Kennedy	to	Full Name: John Kennedy
	Decomposition	Full Name: John Kennedy	to	First Name: John Family Name: Kennedy
	Order change	John Kennedy to Kennedy, John		

Annex A
(informative)

Relationship between metadata mapping and registry

When 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.



NOTE ISO/IEC 11179-6:2005, Figure C-1.[4]

Figure A.1 — Relationship between metadata mapping and registry

Annex B
(informative)

Factoring and types of heterogeneity

B.1 Factoring

B.1.1 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 part of ISO/IEC TR 20943 provides the classes into which data element descriptions are sub-divided; it does not address how one factors of 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 part of ISO/IEC TR 20943 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.1.2 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 family in ISO/IEC 11404[5]).

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

