

# ISO/IEC JTC 1/SC 32 N 1811

Date: 2007-11-20

REPLACES: --

<p style="text-align: center;"><b>ISO/IEC JTC 1/SC 32</b></p> <p style="text-align: center;"><b>Data Management and Interchange</b></p> <p style="text-align: center;"><b>Secretariat: United States of America (ANSI)</b> <b>Administered by Farance Inc. on behalf of ANSI</b></p>
--

<b>DOCUMENT TYPE</b>	National Body contribution
<b>TITLE</b>	ISO/TC171/SC2 N507E - Revised draft ISO/NWI/PDTR 11864, Document management - Guidelines for the creation of a metadata crosswalk
<b>SOURCE</b>	WG2
<b>PROJECT NUMBER</b>	1.32.02.
<b>STATUS</b>	for discussion and action at next WG2 meeting at Vilamoura
<b>REFERENCES</b>	
<b>ACTION ID.</b>	FYI
<b>REQUESTED ACTION</b>	for review by WG2
<b>DUE DATE</b>	
<b>Number of Pages</b>	16
<b>LANGUAGE USED</b>	English
<b>DISTRIBUTION</b>	P & L Members SC Chair WG Conveners and Secretaries

Dr. Timothy Schoechle, Secretary, ISO/IEC JTC 1/SC 32  
Farance Inc \*, 3066 Sixth Street, Boulder, CO, United States of America  
Telephone: +1 303-443-5490; E-mail: [Timothy@Schoechle.org](mailto:Timothy@Schoechle.org)  
available from the JTC 1/SC 32 WebSite <http://www.jtc1sc32.org/>  
\*Farance Inc. administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI



**ISO/TC171/SC2 N 507 E**

Date: 2008-10-14

**ISO/TC171/SC2  
Document Management Applications  
Application Issues  
SECRETARIAT: ANSI**

**TITLE** : Revised draft ISO/NWI/PDTR 11864, Document management – Guidelines for the creation of a metadata crosswalk

**SOURCE** : ISO/TC 171 SC2 Secretariat

**PROJECT** : WI-11864

**STATUS** : Draft

**REQUESTED ACTION** : For discussion in Beijing.

**DISTRIBUTION** : P, 0 and L Members

Address Reply to:  
Secretariat - ISO TC171 SC2- Association for Information and Image Management International  
1100 Wayne Ave, Suite 1100, Silver Spring, MD 20910-5603  
Telephone: 301-755-2682; Facsimile: 240-494-2682; e-mail: [bfanning@aiim.org](mailto:bfanning@aiim.org)

**ISO TC 171/SC 2 NXX**

Date: 2008-10-14

**ISO/PDTR 11864**

ISO TC 171/SC 2/WG 6

Secretariat: ANSI

## **Document management — Guidelines for the creation of a metadata crosswalk**

Document type: Technical Report  
Document subtype:  
Document stage: (20) Preparatory  
Document language: E



### 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:

[Indicate the full address, telephone number, fax number, telex number, and electronic mail address, as appropriate, of the Copyright Manger of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the working document has been prepared.]

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

Violators may be prosecuted.

# Contents

Page

1	Scope .....	5
2	Normative references .....	5
3	Definitions .....	5
4	Metadata Crosswalk properties.....	6
4.1	Crosswalk properties .....	6
4.1.1	Schema Identification.....	6
4.1.2	Categorization .....	6
4.1.3	Crosswalk entry .....	6
4.1.4	Crosswalk administration and help .....	6
4.2	Category and subcategory semantic definition properties.....	7
4.2.1	Controlled vocabulary .....	8
4.2.2	Category and subcategory definitions .....	8
4.2.3	Resource Description Framework (RDF) language.....	8
4.3	Metadata element semantic definition template properties .....	8
4.3.1	Metadata element description template .....	9
4.3.2	Template sample.....	9
5	Crosswalk Functionality .....	11
5.1	Baseline (common) schema .....	11
5.2	Initial crosswalk content .....	11
5.3	Schema reconciliation.....	11
5.4	Crosswalk prototype .....	11
5.5	Links to exiting metadata crosswalks .....	11
6	User functionality and interface .....	11
6.1.1	Schema preparation .....	11
6.1.2	Upload functionality .....	11
6.1.3	Online tools .....	11
6.1.4	Download functionality .....	12
6.1.5	Language capabilities .....	12
6.1.6	Access for Individuals with Disabilities .....	12
7	Crosswalk and semantic definition data structure .....	12

## 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 11864 was prepared by Technical Committee ISO/TC 171, *Document Management Applications*, Subcommittee SC 2, *Application Issues*.

## Introduction

The document management industry has become heavily reliant on diverse metadata. Effective document management often requires the translation between metadata descriptions and schemas. Therefore, a generalized crosswalk is needed between the metadata schemas in common usage in document management applications. This document will provide document management participants with a method and a guideline on how to create a metadata crosswalk.

The implementation of the crosswalk will not only include document management schema but also other applications generating digital files such as records management or

The objective is to define the method by which an online metadata crosswalk may be designed, built and implemented. The crosswalk design will allow owners and developers of metadata schemas to post the elements and semantic definitions of their schemas to the crosswalk. The owners and developers of the schemas will also have the tools available to link the elements of their schema to the element of the common schema and download linked schemas previously posted to the online crosswalk.

## 1 Scope

This document will provide the user with a guide to create an online metadata crosswalk. Metadata elements addressed in this report will include those used in document management processes. Submitted metadata schemas could describe digital or physical objects.

## 2 Normative references

The following document contains provisions which, through reference in this text, constitute provisions of this International Standard. At the time of the publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on the International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below.

*Referenced standards will be listed here.*

ISO 11179 Metadata Depository

(DIS) ISO 13606

## 3 Definitions

Terms are defined in ISO 12651 Electronic Imaging—Vocabulary.

*Additional definitions unique to metadata crosswalks and this document will need to be added in this section*

### 3.1

Information Object

An information object is a digital item or group of items, regardless of type or format that can be addressed or manipulated as a single object by a computer. This concept can be confusing in that it can be used to refer both to actual content (such as digitized images) and to content surrogates (such as catalog records or finding aids). Setting the Stage, By Anne J. Gilliland p 3 in Introduction to Metadata, Pathways to Digital Information, Online Edition, Version 2.1

### 3.2

Schema

Ontology (computer science), a data model that represents the relationships of a set of concepts within a domain (Wikipedia) or a schema in general is a specific, well-documented, and consistent plan.

NOTE XML schema, a way to define the structure, content and, to some extent, the semantics of XML documents  
<http://en.wikipedia.org/wiki/Schema>

### 3.3

Wiki

a collaborative website which can be directly edited by anyone with access to it.

<http://en.wikipedia.org/wiki/Wiki>

### 3.4

Harmonization

the process of enabling consistency across metadata standards. Harmonization of metadata standards is essential to the successful development of crosswalks between metadata standards. Harmonization results in the ability to create and maintain only one set of metadata, and to map the metadata to any number of related metadata standards. The use of harmonization vastly simplifies the development, implementation and deployment of related metadata standards through the use of common terminology, methods and processes. (Issues in Crosswalking Content Metadata Standards, Margaret St. Pierre & William P. LaPlant, Jr.; <http://www.niso.org/press/whitepapers/crsswalk.html>)

### 3.5

#### Crosswalk

a specification for mapping one metadata standard to another.

### 3.6

#### Controlled vocabulary

A prescribed set of consistently used and carefully defined terms (Dublin Core Metadata Glossary, <http://library.csun.edu/mwoodley/dublincoreglossary.html>)

## 4 Metadata Crosswalk properties

### 4.1 Crosswalk properties

#### 4.1.1 Schema Identification

The crosswalk will provide a place for users to access and update as necessary their schema. Users will also be able to browse all schemas that are already registered and linked.

#### 4.1.2 Categorization

Each metadata element will be assigned a category or subcategory and grouped with elements with similar properties, use and attributes. A list of suggested categories may be found in Section 5.2.

#### 4.1.3 Crosswalk entry

A metadata crosswalk user interface will have the functionality to enter or import in an XML format an organization's crosswalk schema, definition and controlled vocabulary. Minimal submission guidelines for an added crosswalk will be created by the metadata crosswalk owner. All new schemas will be linked to the default elements in the metadata crosswalk by the submitter. The submitter of a new schema will have the responsibility to link specific elements in their schema to specific elements in the common schema.

A tool or utility will be provided for down load by the user to allow them to import their schema and definitions into the crosswalk and make element linking changes after importing their schema.

#### 4.1.4 Crosswalk administration and help

##### 4.1.4.1 Public interface

Users of the crosswalk will be required to register before using the crosswalk and/or adding a metadata schema to the crosswalk. The following information will be required for a user to register online for the crosswalk:

Name, email, website, phone number, (contact information), organization

User name and password creation

A help section in the crosswalk will provide contextual help and full guidelines on how to use the crosswalk, upload and download schemas and explain the other functionality of the crosswalk. Within the help section users will be able to provide feedback through a Wiki or some other collaborative tool.

The crosswalk will contain a disclaimer stating the responsibility requirements of users and what the crosswalk will do and what it will not do. Users will need to agree to a user license before entering the crosswalk and using its features.

#### 4.1.4.2 Administrator interface

A crosswalk administrator will allow for the review of schemas submitted and apply the acceptance criteria for any submitted schema. Reviewers will be volunteers that meet a criterion for reviewing metadata schemas. It is suggested the reviewers serve for a set length of time and periodically be replaced.

The crosswalk administrator will maintain users' registration information in a secure format and environment.

#### 4.2 Category and subcategory semantic definition properties

An information object is described by metadata. The features of an information object include content (what an object contains), context (the who, what, why, where and how of an object's creation) and structure (the relationships or associations within or among information objects). Structure in a metadata schema is achieved by categorizing each element with the schema. Table 1 based on Setting the Stage, By Anne J. Gilliland, in Introduction to Metadata, Pathways to Digital Information, Online Edition, Version 2.1, p3 provide the default categories suggested for a metadata crosswalk.

**Table 1 — Different Types of Metadata and their Functions**

Type	Definition	Examples
<b>Administrative</b>	Metadata used in managing and administering information resources	- Acquisition information
		- Rights and reproduction tracking
		- Documentation of legal access requirements
		- Location information
		- Selection criteria for digitization
		- Version control and differentiation between similar information objects
		- Audit trails created by recordkeeping systems
<b>Descriptive</b>	Metadata used to describe or identify information resources  (Indexing language)  (Control language to describe a field)	- Cataloging or describing records
		- Finding aids
		- Specialized indexes
		- Hyperlinked relationships between resources
		- Annotations by users
		- Metadata for recordkeeping systems generated by records creators
<b>Preservation</b>	Metadata related to the preservation management of information resources	- Documentation of physical condition of resources
		- Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration
<b>Technical</b>	Metadata related to how a	- Hardware and software documentation

	system functions or metadata behave	- Digitization information, e.g., formats, compression ratios, scaling routines
		- Tracking of system response times
		- Authentication and security data, e.g., encryption keys, passwords
<b>Use</b>	Metadata related to the level and type of use of the resources	- Exhibit records - Use and user tracking- Content re-use and multi-versioning information

#### 4.2.1 Controlled vocabulary

A controlled vocabulary should be used to describe metadata types. Controlled vocabularies are hierarchical in nature going from a broad term to a narrow term for a given category or item. Conducting searches within a metadata crosswalk can be enhanced by searching keywords that are not part of the controlled metadata vocabulary.

Controlled vocabularies consist of organized lists of words or phrases used to provide consistent tagging of content for resource discovery. Taxonomies, thesauri, classification, facets, synonym rings, etc are all controlled vocabularies with some level of controlled language terminology. Strong preference should be given to using a thesaurus for choosing subjects as this will provide consistency in the use of terms and assist with focused searching. (NLA 2002). The DC [Dublin Core] standard also recommends use of a controlled vocabulary or formal classification scheme as best practice (Hillmann, 2005, Using Dublin Core)

Controlled vocabularies can be derived from an existing controlled vocabulary, created specifically for the intended application or a combination of both approaches. An example of a controlled vocabulary for a document management metadata crosswalk is Simple Knowledge Organisation System (SKOS).

The controlled vocabulary for a metadata schema, if it exists, will also be imported into the metadata crosswalk along with the schema itself.

#### 4.2.2 Category and subcategory definitions

Category and subcategory definitions will be available for users to categorize their metadata elements based on their function (category: Field of interest, e.g. aerospace)

#### 4.2.3 Resource Description Framework (RDF) language

The metadata crosswalk should use Resource Description Framework (RDF) language in its development. The Resource Description Framework (RDF) integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web. (<http://www.w3.org/RDF/>) The purpose of RDF is to create machine-processable statements. This can be accomplished by using the OWL Web Ontology Language (<http://www.w3.org/TR/owl-features/>). OWL's purpose is to present information in a format that can be processed by applications, as opposed to situations where the content only needs to be presented to humans.

#### 4.3 Metadata element semantic definition template properties

The metadata semantic definition template should use Resource Description Framework (RDF) terminology in constructing the template. Please see section 4.2.3 for a description of RDF language

#### 4.3.1 Metadata element description template

Sections of template should be included as part of the schema registration.

#### 4.3.2 Template sample

A sample template is found below in this document. Each entity posting a metadata schema to the online crosswalk would be required to complete the semantic definition template/form for each metadata element contained in their schema.

**Table 2 — Possible data elements for semantic definition template**

Element Name	A unique Element, Name, or Identifier.
Description of Element	A description of the unique Element, Name, and if applicable, a range of values
SUB-ELEMENTS	These are called attributes in other Recommended Practice's. They are elements that help to further describe the Metadata Element, Name or Identifier. Include the RDF path to elements. Inheritance rules for elements should be included.
XML CODE	This is the XML Code for the Metadata Element, Name or Identifier and the Sub-Elements. (schema or DTD)
CROSSWALK DEFINITION	In the creation of these Metadata Element there was a review of element found in other Metadata Recommended Practices. Some of these Practices were the Dublin Core, Pandora, Cedar, RLG, ISAD. A list of these elements and their definitions are provided in the Crosswalk. These elements have assisted in the creation of the elements and definition which are contained in this Recommended Practice.
Version	Allowing any metadata element description to be edited and revised while retaining the history of the changes such that an older version of the file can be recalled, if necessary. Files will retain information about the original sources, so that the resulting edited media can be traced back to its original source. Version number and Dates will be required.
Registration Authority	The user fills out this area to identify the authorized agency or individual that is responsible for registering the data element.
Reason for Data	This provides information regarding the condition upon which the Rationale for the data was based. How the metadata will be used in the document management process. Why do we have it!
Language	The language of the content of the data objects being archived. The language in which the information is presented, e.g. English, German, French, etc. (use ISO 639 standard for 2-letter acronym for languages)
Origin of Metadata	The area from which the metadata was originally obtained. (e.g., Camera Settings from the Camera, Publisher from the Archive, Date from Record, date of capture, etc.) (activity or industry that uses the schema)
Nature of metadata	Lay metadata created by persons who are neither subject nor information specialists, often the original creator of the information object Expert metadata created by either subject or information specialists, often not the original creator of the information object

Status	<p>Static metadata that never change once they have been created</p> <p>Dynamic metadata that may change with use or manipulation of an information object</p> <p>Long-term metadata necessary to ensure that the information object continues to be accessible and usable</p> <p>Short-term metadata, mainly of a transactional nature</p>
Structure	<p>Structured metadata that conform to a predictable standardized or non-standardized structure</p> <p>Unstructured metadata that do not conform to a predictable structure</p> <p>(Formalized vs. non-formalized schema)</p>
Instances	Number of instances permitted of each element: one or multiple
Semantics	<p>Controlled metadata that conform to a standardized vocabulary or authority form</p> <p>Uncontrolled metadata that do not conform to any standardized vocabulary or authority form</p>
Level	<p>Collection metadata relating to collections of information objects</p> <p>Item metadata relating to individual information objects, often contained within collections</p>
Obligation <b>(not in the common schema)</b>	<p>Terms such as Mandatory, Conditional and Optional have been avoided. They are so closely associated with data input models.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> <b>Essential information (Mandatory):</b> will definitely be required.</li> <li><input type="checkbox"/> <b>Essential if applicable (Conditional):</b> refers to the fact that some elements are more relevant to some materials or processes than others, so they may be essential if applicable.</li> <li><input type="checkbox"/> <b>Desirable Information (Optional):</b> will not be critical, but is expected to be helpful.</li> </ul>
Comments	Additional comments as needed to further help define the Metadata Element, Name, or Identifier.

## **5 Crosswalk Functionality**

### **5.1 Baseline (common) schema**

A baseline schema with categorized elements will be created in the crosswalk. All other schemas will be linked to corresponding elements in the baseline schema by the submitter. An uploaded schema need not be linked to all the elements in the baseline schema. A unique identification number will be assigned to each element in the baseline schema. Posted schemas will each have a unique identifier as will the elements in the schema. This approach will ensure that all schemas will be linked to each other. There will be versioning control for the baseline schema to track any changes to the content.

A user or submitter may request that a new category, subcategory or element be added to the category list or to the baseline schema. A committee of reviewers will approve or deny such requests.

### **5.2 Initial crosswalk content**

The initial crosswalk will start with a set grouping of registered schemas that are common to the domain of document management. It is suggested that existing schemas for document management in the specific industry be used to populate the initial schema.

### **5.3 Schema reconciliation**

Harmonization between schemas (list of fields) and terminology reconciliation should be automated as much as possible with minimal human intervention.

### **5.4 Crosswalk prototype**

It is suggested that a prototype of the intended crosswalk be developed to help clarify the scope of the online repository, its scalability, performance and the number of simultaneous users expected to use the crosswalk.

### **5.5 Links to exiting metadata crosswalks**

Links to existing metadata crosswalks could be included in the online crosswalk. These links could be posted by the administrator or submitted by users.

## **6 User functionality and interface**

### **6.1.1 Schema preparation**

Downloadable instructions will be available to help users prepare their schemas for inclusion in the metadata crosswalk. A tool may also be downloaded for users to format their schema and semantic definitions, add controlled vocabulary if available and link their schema to the baseline schema. The baseline schema will be included in the linking tool. The output from this tool could automatically be uploaded and ingested into the crosswalk (system).

### **6.1.2 Upload functionality**

A submitter will be able to upload the output from the schema linking tool directly into the online metadata crosswalk. Some submitted crosswalk elements might not be bi-directional if both schemas are not linked to the same baseline elements.

### **6.1.3 Online tools**

An online interface will provide a submitter access to their schemas to verify the linking and make changes as needed.

If a submitter feels that an element is missing from the baseline set, he or she can make an online request that an element be added to the baseline schema. The submitter will need to justify adding the element to the baseline schema. The element will be reviewed by a crosswalk review committee and added if approved. Otherwise the submitter will need to link the element to another element, link it to a category or subcategory or tag it as an orphan.

Suggest how often the committee reviews a new element to the baseline schema. Need criteria for adding an element to the baseline schema. Should new elements be labeled as candidates until review takes place?

A user will be able to view other schemas and elements that are linked to their schema posted on the online crosswalk. By clicking on an element the user will be able to see the definition and other information about the element.

A submitter will have the tools to update their schema online. There will be versioning control for schemas posted to the crosswalk.

#### **6.1.4 Download functionality**

The user interface to the crosswalk will allow a user to download linked schemas in XML format with a style sheet that could contain documentation about categories and subcategories. Schema element definitions will also be available for download along with its corresponding schema. The user will be able to download the following files:

Schema, documentation, semantic definitions, controlled vocabulary and tag library

A user will be able to view and select which linked schemas they wish to download in an XML format from a list of schemas included in the crosswalk using a dropdown box or some other listing tool.

#### **6.1.5 Language capabilities**

A crosswalk should provide a user interface with various language options.

#### **6.1.6 Access for Individuals with Disabilities**

Consideration should be made in developing a crosswalk for those individuals with disabilities that need special accommodations to use the crosswalk. Please refer to ISO 9241 Ergonomic Office work standard.

### **7 Crosswalk and semantic definition data structure**

The download data structure for the crosswalk will be in a XML format.

The crosswalk developer will decide on the data structure based on functional requirements as expressed in this document. An explanation about the hierarchical structure of a crosswalk plus other germane data structure information will be provided to crosswalk users through downloading a file with the information.

The crosswalk creator/owner will provide an outline for data structure for the elements including header information, unique identifiers for each crosswalk created, file size for schema and creation information.

The data structure for the schema could be created in a tree structure. An example of such a structure can be found in the Encoded Archival Description DTD (<http://www.loc.gov/ead/>).