

**ISO/IEC JTC 1/SC 34**

Date: 2005-02-12

**ISO/IEC CD 13250-5**

ISO/IEC JTC 1/SC 34/WG 3

Secretariat: SCC

## **Topic Maps — Part 5: Reference Model**

### **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 document 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 Draft International Standard and is copyright-protected by ISO. Except as permitted under the applicable laws of the user's country, neither this ISO draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured.

Requests for permission to reproduce should be addressed to either ISO at the address below or 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 may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

**Contents**

Page

Foreword.....	iii
Introduction.....	iv
1 <b>Scope</b> .....	1
2 <b>Glossary</b> .....	1
3 <b>Topic Map Applications (TMAs)</b> .....	2
4 <b>Conformance of Disclosures of Topic Map Applications</b> .....	2
5 <b>Formal Model</b> .....	3
<b>Annex A (informative) Useful References</b> .....	4

## 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.

ISO/IEC 13250-5 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 34, Document Description and Processing Languages.

ISO/IEC 13250 consists of the following parts, under the general title *Topic Maps*:

- *Part 1: Overview and Basic Concepts*
- *Part 2: Data Model*
- *Part 3: XML Syntax*
- *Part 4: Canonicalization*
- *Part 5: Reference Model*

## Introduction

Anyone who has played the game of "twenty questions" already has a basic understanding of the principles of topic maps. When this game is played, one person thinks of a person, place or thing and the other players ask questions which must have yes or no answers. If a player has an idea as to the identity of the person, place, or thing, based upon answers given, that player can name it. If the idea turns out to be correct, the insightful player becomes the next person to think of a person, place or thing.

In the terminology of the Topic Maps Reference Model (TMRM) (this International Standard), the leader of the game is thinking of a "subject". As the game progresses, the players ask questions such as: "Is it a person?", "Is the person alive?", "Is the person female?", "Is she a rock star?", etc. The responses to those questions are used by the players to guess the subject the leader has in mind.

Successful participation in a game of twenty questions depends upon the sharing of several unspoken assumptions. Those unspoken assumptions are similar to the assumptions that must be made explicit in order to merge subject proxies (representatives of subjects) uniformly across information systems. The purpose of the TMRM is to provide a checklist of the assumptions that must be made explicit in order to support such uniformity.

In any game of twenty questions, there must be agreement among the players that the only questions that will be asked are questions that can be answered "Yes" or "No". One way of making this underlying assumption explicit is to say that subjects will be identified by means of some set of twenty or fewer yes-or-no questions, together with their yes-or-no answers. Nobody would claim that this is always the most efficient or effective way to identify subjects (twenty questions is just a game, after all), but it certainly is one way to establish the identity of a subject. In the much broader context of the world's diverse information resources, subjects and ways of identifying subjects are constantly being invented. Every approach to the problem of subject identification offers both benefits and limitations; different subjects and different motivations demand different approaches.

Players of the game of twenty questions do not normally stop to discuss the various interpretations that can be made of any question or its answer. If the game can be played at all, the players share enough language, culture and experience to understand the significance of each question and answer. In the broader context of the world's information resources, too, indications of subjects are recognized only within limited ontological, cultural and/or technological contexts. If we want to gather all the information about a given subject, but the ways in which the subject is indicated are ontologically, culturally, and/or technologically diverse, our task is only possible if we know what those ways are, and how they work.

There is no shortage of proposed standard ontologies or terminologies competing for the status of becoming a dominant ontological context for the representation of subjects in various areas. The utility of many such approaches rests upon two questionable assumptions: First, that any standard ontology can represent the full diversity of past, present or future ontologies that underlie the subjects that various persons may perceive to be indicated in information resources. Second, that whatever information losses may occur in the transposition to a standard ontology are acceptable.

The TMRM requires that each disclosure of a Topic Map Application (TMA) specify the ontological assumptions that are made in order to indicate a subject within an ontology. Choosing any particular structure, syntax or data model for TMAs (or topic maps) would of necessity exclude or constrain ontologies and/or means of indicating a subject within an ontology, and so the TMRM focuses solely on disclosure of those choices.

Appropriate standards bodies will create standard TMAs. That is expected and necessary to serve the needs of various communities. The disclosure of the ontological choices made in those TMAs will enable users across diverse communities to effectively marshal information about subjects indicated differently in diverse communities.

TMAs govern the topic maps with which they are associated. Disclosures of TMAs are constrained by the TMRM in order to facilitate the sharing of understandings about structure of and rules governing topic maps and indications of subjects in topic maps. The TMRM expresses those constraints by defining subject proxy as an abstraction, with no express or implied limitations on the structure or structures of subject proxies -- the structure(s) in terms of which a TMA is disclosed.

The disclosure required for conformance to the TMRM enables diverse information resources to be viewed through the ontological assumptions of a single TMA, and for information resources to be viewed through the ontological assumptions of multiple TMAs, and for views of diverse resources through diverse TMAs. All resources can be seen as sets of subject proxies. Rather than opposing a diversity of viewpoints, the TMRM embraces them and facilitates their integration, thus enabling richer, more inclusive and more useful views of the world.

# Topic Maps — Part 5: Reference Model

## 1 Scope

This International Standard specifies:

1. The abstract definition of the term "subject proxy."
2. The abstract definition of the term "Topic Map Application" (TMA), and requirements to be met by disclosures of TMAs.
3. Other definitions and specifications in support of the above.

This International Standard does not specify:

1. The subjects that may be indicated by subject proxies or constraints on subjects or subject proxies.
2. The algorithms, data models or syntaxes that may be used to represent, generate, view, compare, or process subject proxies.

## 2 Glossary

### Property

A value appearing in a subject proxy that is governed by a TMA.

NOTE 1 How a property comes to appear in a subject proxy is deliberately unspecified. Properties may appear by any means, including but not limited to physical entry of the property, auto-generation by any means, or the operation of any TMA-disclosed rule, or any combination of such means.

To avoid constraining the designs of TMAs or their implementations, properties are not required to have names. However, TMAs are required to identify the properties they govern. TMAs must specify how that is accomplished.

### Subject

Any thing whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be indicated by any means whatsoever.

### Subject Proxy

A unit of information that is a set of one or more properties, each of which is governed by a specific TMA, and that, in the light of its governing TMA, indicates a single subject.

NOTE 2 The Topic Maps Data Model, ISO 13250-2, can be disclosed as a Topic Map Application. It defines the structure of "topic", "association", and "occurrence" subject proxies, along with specific rules and requirements for their use. The term "subject proxy" refers to an abstraction used herein to describe specify the disclosures required for interchange of topic maps governed by the Topic Maps Data Model or any other Topic Map Applications.

### Subject proxy class

An abstract model of the structure of some set of subject proxies. Such a structure provides for the encapsulation of properties within the proxy, for identifying each property in the terms of its governing TMA, and for specifying its value.

NOTE 3 It is the quality of possessing defined properties that in whole or in part (or, in the case of multi-TMA subject proxies, in multiple parts) indicates a subject that defines a subject proxy class.

NOTE 4 This International Standard does not define any subject proxy classes.

NOTE 5 A TMA may define multiple subject proxy classes, perhaps with the expectation that certain kinds of subjects will be indicated by instances of certain classes of subject proxy; this is the approach used in the Topic Map Data Model (ISO 13250-2). A second alternative is that a TMA may define a single class of subject proxy that has been designed in such a way that its instances can indicate any of the subjects contemplated by the TMA. A third alternative is that a single class of subject proxy can be designed to be used to indicate all of the subjects contemplated by any number of independent TMAs simultaneously; this approach can be used to allow subject proxy governed by different TMAs to be merged to form subject proxies whose various parts are governed by their respective TMAs.

### Topic Map

A set of subject proxies as defined and governed by one or more Topic Map Applications.

#### Ed. Note.

This definition makes it clear that what qualifies as a topic map is wholly within the domain of one or more Topic Map Applications and their disclosures. That is to say that the TMRM defines the general characteristics of a topic map but those characteristics are realized only through the mechanism of a TMA.

It also avoids the use of the possibly confusing phrase "topic map views." There are only topic maps as defined by their TMAs.

### Topic Map Application (TMA)

A set of constraints, disclosed in conformance with the requirements of this International Standard, on the structure, processing, and interpretation of the properties of Subject Proxies.

## 3 Topic Map Applications (TMAs)

A Topic Map Application (TMA) must disclose the following for each subject proxy class governed by that TMA:

1. The structure of the subject proxy class.
2. The properties defined for the subject proxy class.
3. Rules for comparison of instances of subject proxies, based upon the properties disclosed as indicating a subject by the governing TMA, to determine whether the subject proxies indicate the same subject.
4. Rules for viewing multiple subject proxies that have been deemed to represent the same subject as a single subject proxy.

NOTE In some cases, the subject being indicated by a subject proxy may be ambiguous or open to conflicting interpretations, even when the disclosure of the applicable rule for determining whether two subject indications are equivalent is unambiguous. While it is not possible to avoid all ambiguity or potential conflicting interpretations, the usefulness of any TMA will depend upon the extent to which ambiguities or conflicting interpretations concerning the subjects being indicated by subject proxies are avoided.

5. Rules, if there are any, for viewing the topic map in such a way that properties whose existence in the topic map is implicit become explicit.
6. Rules for identification of a subject proxy with its governing TMA, and for associating each property of each class of subject proxy with whatever portions of the TMA disclosure are applicable to that component property. If the properties of a subject proxy may be able to be governed by different TMAs, the rules for specifying which properties are governed by which TMAs must also be disclosed.

## 4 Conformance of Disclosures of Topic Map Applications

If a disclosure of a Topic Map Application meets the disclosure requirements of section 3, then it is a conforming Topic Map Application Disclosure.

## 5 Formal Model

### **Ed. Note.**

Due to time constraints on the author of the Tau model (Robert Barta) and the editors, the formal model based upon the Tau model is not included. After studying the latest version of the Tau model, with written and verbal communication with Robert Barta, the editors believe that the TMRM is consistent with the Tau model and its formalism. An editor's draft will be prepared before the next meeting of WG3 that incorporates that formalism.

## Annex A (informative)

### Useful References

The problem of "subject identity" has recently been recognized as more difficult than previously thought by proponents of the Semantic Web. The following annotated references illustrate this problem and survey various approaches.

Croucher, Tom and Joe Geldart. Situation and Identity: A Generalization of Inverse Functional Properties. (publication pending, <http://osiris.sunderland.ac.uk/~cs0tco/eswc2005.pdf>)

Croucher and Geldart note three forms of identity, lexicographic (unique names or PSIs), denotational (different indicators for the same subject), and concept identity. This paper focuses on use of a subsumption hierarchy to address the issue of denotational identity.

Of particular interest is the authors' statement: "We suggest that until concept identity has been properly defined the semantic web will be missing a core tool for identification of objects." Section 6

From the TMRM perspective, concepts are simply subjects, each of which can be indicated (denoted) via properties of subject proxies. The question of concept identity is therefore not separate from denotational identity but a question of how to denote concepts in an information resource.

Guha, R. Object Co-identification on the Semantic Web, WWW2004, May 17-22, 2004, New York, NY USA.

Guha characterizes reliance on standardized ontologies providing common names (or URIs) for the Semantic Web as "overly optimistic," later commenting that "it will be almost impossible to get agreement on URIs for all people, places, products, ...which is why an alternative approach is required for these."

Oddly, he then proceeds to suggest that agreement would be possible on "names for schema level objects such as property types and classes, but not for all individual objects." Based on that supposed bootstrapping, Guha introduces the notion of Discriminant Descriptions (DD) which rely upon the DD being a discriminant across the union of two domains. Using structured data for cities across the world and employees of IBM in the US, and probabilistic matching, he was able to achieve 99%+ accuracy in matching between the two data sources.

Finally, Guha cites related and older literature on "record linkage," data integration and co-identification of people across data sources. Many of the "new" problems encountered by the Semantic Web are in fact well known and addressing them will proceed more quickly if prior work is not ignored.

Sheth, Amit, Cartic Ramakrishnan, Christopher Thomas, Semantics for the Semantic Web: The Implicit, the Formal and the Powerful. Int'l Journal on Semantic Web & Information Systems, 1(1), 1-18, Jan-March 2005.

The authors divide approaches semantics into implicit (patterns in data but not explicit), formal (explicit, description logic, OWL), and powerful (statistical analysis, probabilistic reasoning). It is an extensive survey of current approaches with a bibliography of almost three pages.

Given the current popularity of description logic and its being the underpinning for OWL, it is evaluated closely by the authors. The problems with description logic are discussed both under the treatment of description logic but also in contrast to the other semantic approaches. The authors note that the Semantic Web should not be limited to one type of representation at the expense of others.

Each of the approaches suggested by this article, could in fact be understood and disclosed as a TMA, which would facilitate the exploitation of the resulting subject proxies in the contexts of other subject proxies -- subject proxies whose properties may be governed by the same or different TMAs.