# Building ISO Topic Maps in OWL-DL

**Anne Cregan, PhD student**
annec@cse.unsw.edu.au

*Knowledge Representation and Reasoning Program,*
*National Information and Communication Technology*
*Australia (NICTA) Centre of Excellence*

*Artificial Intelligence Group*
*School of Computer Science and Engineering,*
*University of New South Wales, AUSTRALIA*

## 1. The Topic Map Tradition

Topic Maps are an ISO/IEC standard **[TM 02]** for mapping web and "real-world" information resources, by reifying real-world resources as "subjects", and creating "topic" constructs to capture their characteristics and relationships with other topics and subjects. Dubbed the "GPS of the information universe", they are akin to an electronic "back-of-book" index, supporting information navigation and retrieval in on-line environments. In 2001, two ISO 13250 standard syntaxes for Topic Maps were established, one in HyTM and another in XML **[TM 02]**.

## 2. The Semantic Web Initiative

The Semantic Web is a W3C-led initiative with the goal of providing technologies and standards for the semantic markup of information resources, thus enabling improved web navigation and supporting intelligent web services. Like XTM, it is also an XML-based technology, using Resource Description Framework (RDF) and Web Ontology Language (OWL) layers superimposed on XML to provide more expressive representations of the characteristics of, and relationships between, logical entities. OWL-DL, a formal W3C recommendation finalised in 2004, is a subset of OWL-Full which has a description logic equivalent semantics.

## 3. Topic Map/RDF Interoperability Problems

Although the Topic Map standards were developed independently of the W3C's Semantic Web initiatives, it has long been suspected that as both Topic Maps and the Semantic Web have the same goal, namely to be meta-level maps of information entities, there must be synergies to be exploited. However, previous attempts at achieving interoperability between Topic Maps and RDF have been problematic for the following reasons, amongst others:

- Neither Topic Maps nor RDF had stable formalized data models.
- Topic Maps use n-ary relations, and both direct and indirect locators. RDF is not sufficiently expressive to represent this.
- The use of Topics as Types within Topic Maps has been problematic – should they be interpreted as classes, instances, or both?

A full analysis is available in a survey of Interoperability Proposals recently published by the W3C's RDF/Topic Maps Interoperability Task Force **[RDFTM 04]**.

## 4. The Topic Map Data Model (TMDM)

The original Topic Map standard did not explain the relation between its two syntaxes, and did not make the common underlying data model explicit. Subsequently drafted requirements for a Topic Map Constraint Language and Query Language both require a clear description as to how Topic Map constructs are to be evaluated. Accordingly, the ISO Topic Map Working Group commenced work on a Topic Map Data Model in May 2001, culminating in the official draft ISO 13250-2 TMDM **[TMDM 05]** published in January 2005, which irons out some of the problems with the previous representation and gives Topic Maps a formal data model.

The proposed TMDM offers some significant benefits over previous Topic Map specifications. It is represented as an Entity-Relationship diagram, explicitly stating the directions and cardinalities of relationships between Topic map entities. It states type and parent relationships separately and explicitly. It clearly represents the different kinds of relationships Topics may have with Subjects. It gives data attributes for its logical entities, which include values, data types and locators such as URI references. It also introduces two additional constraints to ensure the consistency of the resulting Topic Maps.

## 5. Web Ontology Language (OWL)

OWL has been designed for the representation of classes containing individuals, which may be related via a class-subclass hierarchy. Individuals may be related to each other via OWL's Object Properties, and may also have data attributes, represented by OWL's Data Properties. Both Object and Data Properties may be constrained on domain, range, cardinality and properties such as transitivity. Although OWL properties are one-directional, they may be linked to a corresponding property in the opposite direction by using OWL's "inverse" relation construct. OWL is therefore an excellent candidate to represent an Entity-Relationship model such as the TMDM.

## 6. Representing the TMDM in OWL-DL

The ISO's drafting of an explicit Topic Map Data Model (TMDM) **[TMDM 05]**, combined with the advent of the W3C's Description Logic-equivalent Web Ontology Language (OWL-DL) **[OWLDL 04]** together provide the means for the construction of a Description-Logic equivalent, unambiguous semantic model for Topic Maps.

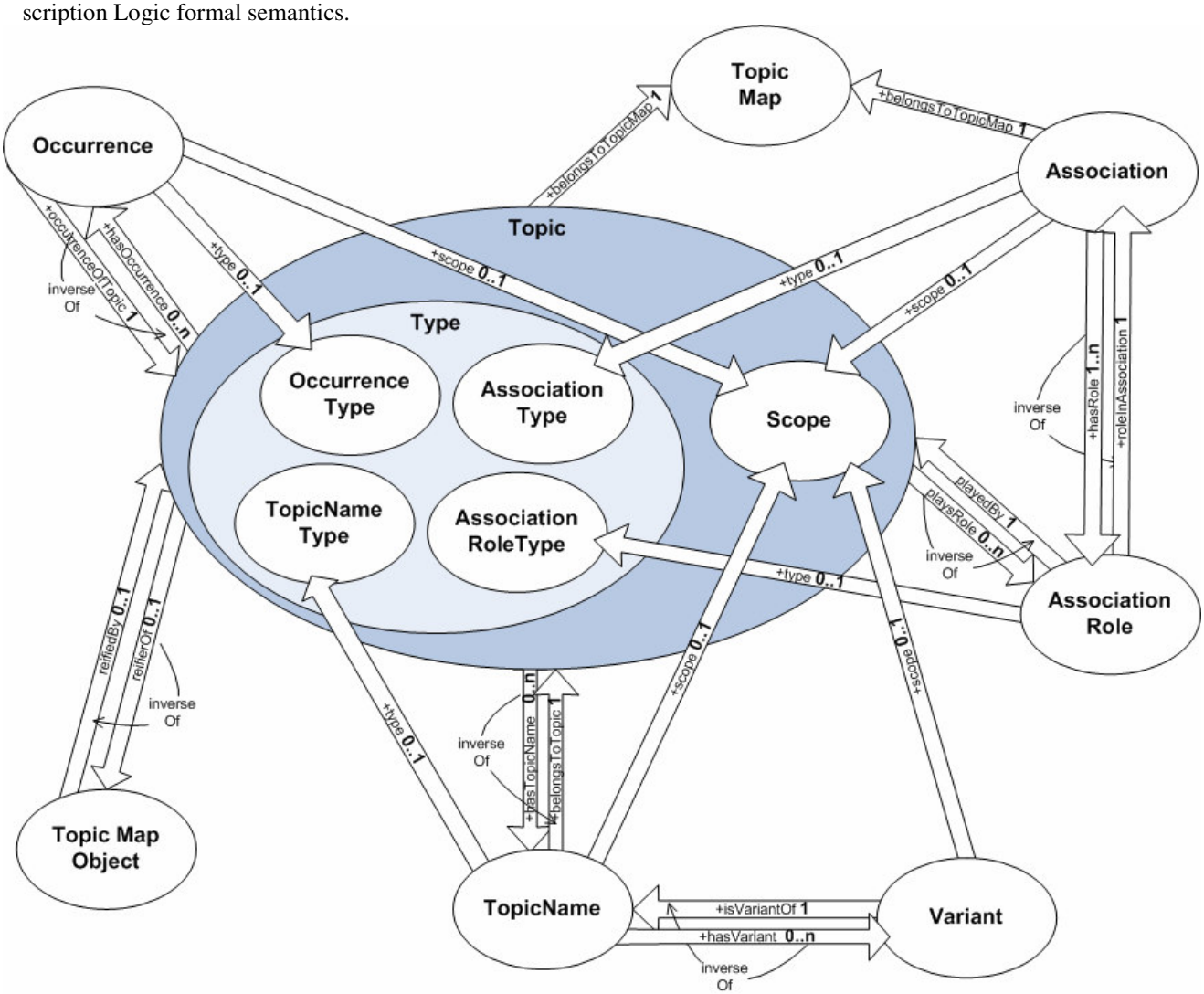
The proposed TMDM has been built as an OWL-DL ontology, using Protégé **[Protege]**, The ontology's classes and object properties are represented in the figure below. Once such a TMDM construction in OWL is ratified & published, it may be used by Topic Map authors to construct their own Topic Maps in OWL using the TMDM constructs. Sample Topic Maps have been built in Protégé/OWL by importing the TMDM ontology.

The advantages of using an OWL-DL construction for the TMDM over the existing standards are:

- OWL allows constraints to be explicitly stated;
- OWL has a mature suite of tools, such as Protégé combined with RACER, which provide the means for querying and enforcing constraints, meeting many of the ISO's requirements for a Topic Map Query Language and a Topic Map Constraint Language
- OWL-DL and OWL-Lite have a clearly defined Description Logic formal semantics.

The advantages to the end-user of building Topic Maps as OWL ontologies which import the TMDM ontology are:

- The TMDM provides Topic Map Constructs which allow users to create their own Topic Maps in OWL, which conform with the standards set by the ISO Topic map committee.
- The user may draw on OWL's constraint tools & capabilities, such as those provided by Protégé/RACER, to support validity and consistency checking, both to enforce the standard TMDM semantics, and to implement additional user-defined constraints.
- Provides access to OWL's suite of APIs, querying tools, visualisation capabilities, etc, such as those Protégé provides.
- Enables the construction of user-defined Topic Maps which have a formal Description Logic semantics.
- Supports the conversion of any existing Topic Map which fits the TMDM into OWL-DL.




**OWL-DL Topic Map Data Model – Graphical Representation**

## 7. Ontology Design Choices

OWL comes in three flavours with increasing degrees of expressivity and complexity: OWL-Lite, OWL-DL and OWL-Full. OWL-Full is the most expressive but also makes reasoning undecidable so is to be avoided if possible. Both OWL-Lite and OWL-DL are decidable, and have a formal Description Logic semantics. OWL-Lite may ultimately prove to be sufficiently expressive to represent the full TMDM, but as the question is still open, OWL-DL is presented as the representation language of choice. Users retain the option to build OWL-Full Topic Maps based on OWL-DL Topic Map constructs.

As with the construction of any ontology, some design choices must be made. The first task is to identify the nature of the individuals in user Topic Maps, recalling that whilst OWL's Object and Data Properties are defined with classes as their domains and ranges, the properties belong to the individuals within those classes. Accordingly, the candidates for individuals are taken to be those individual Topics, Occurrences, Associations and so on that Topic Map authors would ultimately want to define relationships between, and characteristics for.

Having decided that Topics, Occurrences, Associations and so on are to be individuals, the classes "Topic", "Occurrence" and "Association" and so on are created to contain these individuals. These are then used as the domains and ranges to define OWL Object Properties between individuals, providing the general framework and constraints defining which entities may be related to which other entities, and in what way they relate (cardinality, etc). The Topic Map builder will ultimately instantiate these imported classes and relations with her own Topics, Associations, etc to build her own Topic Maps.

A distinguishing feature of topic maps is that Topics are used as Types, whilst also being used in all the ways that regular Topics are. Normally in an ontology, a type would be represented as a class of individuals. However, if this were to be done here, Topics would need to be represented as both classes and individuals, requiring the use of OWL-Full. thus making reasoning undecidable. As OWL provides the expressivity to define complex relations between individuals using Object Properties, it is preferable to represent Types as individual Topics, and to represent the "type" relation as an Object Property between Topic Map Objects and individual Topics.

## 8. Ontology Construction

Each entity in the TMDM E-R diagram was represented as a class in the TMDM ontology (shown as an oval in the **Figure).** Relationships between the entities were captured as Object Properties. Where there were bi-directional relations in the E-R diagram, two Object Properties were defined and related as inverses. For additional ease of use, many additional inverse relations were also defined. Logically, these are already implied by the TMDM: by making them explicit, users are given more flexibility to construct their own Topic Map ontologies in the most convenient way. The attributes of each entity in the TMDM E-R diagram were translated to OWL Data Properties. Constraints were implemented as restrictions on the OWL classes. Sample Topic Maps were then constructed as additional ontologies which imported the TMDM ontology, and added instances representing actual Topics, Occurrences, Associations for specific domains. Note that no XTM references are necessary, either in the TMDM ontology or in user-defined OWL Topic Maps.

Construction in Protégé provided access to Protege's querying tools, constraint checking using RACER, and ontology visualization using Protégé visualization plugins such as Jambalaya. [Protege

## 9. Conclusions and Next Steps

The use of OWL and its associated tools for constructing Topic Maps provides significant advantages over previous Topic Map representations in terms of explicit specification, formal semantics, constraint checking and querying capabilities.

The intention is that once the definitive representation of the Topic Map Data Model in OWL-DL syntax is finalised and approved by the relevant Standards Bodies, it will be published as an online document, available for inclusion by import in User-Defined Topic Maps written in OWL, thus making the TMDM constructs available for building user TMs in OWL. The full OWL-DL construction for the TMDM is currently under review by the ISO TMDM **[TMDM 05]** authors.

## References


**[OWL 04]** *OWL Web Ontology Language Reference,* World-Wide Web Consortium (W3C) Recommendation, 10 February 2004, available at http://www.w3.org/TR/2004/REC-owl-ref-20040210/

**[Protégé]** *Protege Ontology Editor (Version 3.1 beta),* Stanford University open-source software, available at http://protege.stanford.edu/

**[RDFTM 04]** *RDFTM: Survey of Interoperability Proposals,* World-Wide Web Consortium (W3C) RDF/Topic Maps Interoperability Task Force publication, Editors Draft, 24 February 2005, available at http://www.w3.org/2001/sw/BestPractices/RDFTM/survey

**[TMDM 05]** *Topic Maps – Part 2: Data Model* Draft ISO/IEC International Standard (ISO/IEC JTC 1/SC34 ISO/IEC FCD 13250-2), 10 January 2005, available at http://www.isotopicmaps.org/sam/sam-model/

**[TM 02]** *Topic Maps Second Edition*, ISO/IEC International Standard (ISO/IEC 13250*)*, 19 May 2002, available at http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf