

How can XML Schemas Enhance Topic Maps?

Martin **Bryan** <mtbryan@diffuse.org>

Abstract

This paper shows the advantages that can be gained from maximising the use of facilities provided by W3C XML Schemas and the XML Linking Language when defining an XML representation of ISO/IEC 13250 Topic Maps. In particular it highlights the advantages of being able to utilize built-in features of advanced web tools to manage and process topic maps.

The paper also summarizes the way in which Topic Maps have been used to enhance the functionality of the Diffuse project, which aims to inform Information Society systems developers of currently available standards and specifications and their relevance to advanced pan-European R&D projects that are being co-funded by the European Commission

1. What are Topic Maps?

ISO/IEC 13250 defines a set of SGML Architectural Forms for the definition of Topic Maps. These architectural forms are specifically designed to create customized sets of elements for the creation of navigable topic maps based on the advanced linking techniques provided in the ISO/IEC 10744, the [Hypermedia/Time-based Structuring Language (HyTime)].

[Figure 1](#) shows the main components of a topic map:

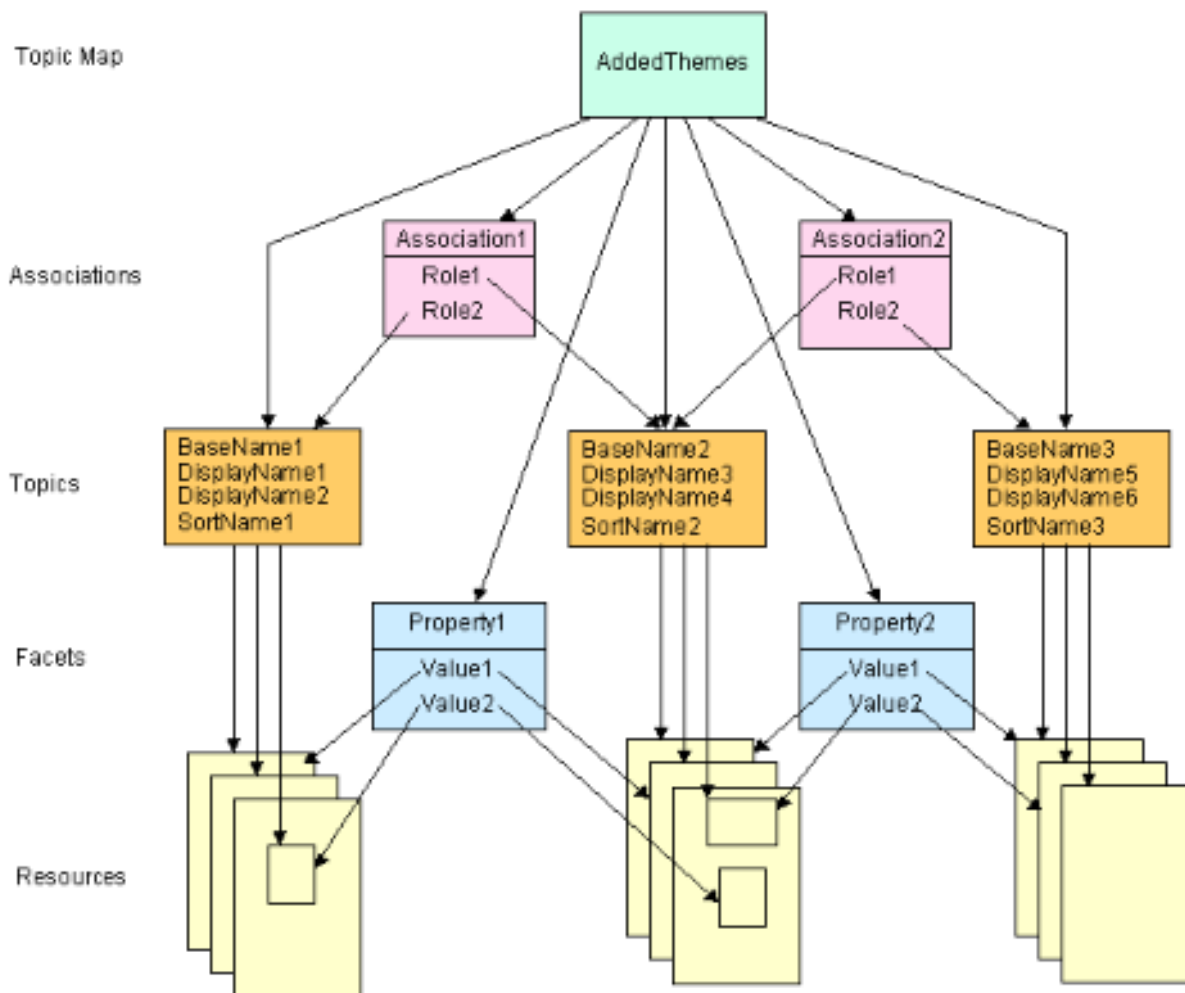


Figure 1. Figure 1

Relationships between elements in topic maps

The topicmap element type is used to identify those parts of a resource that define a navigable topic map. It contains details of individual topics, associations links between topics, facets that assign property values to resource occurrences, and details of themes (topics) that are to be considered to be associated with certain types of topic map components.

A topic consists of a set of topic names that can be used to identify the meaning of a topic, and a set of pointers to the parts of resources that have been identified as using, or explaining, the meaning. In addition to a text-only basename, a topic can be

assigned any number of displayable names, which can be defined using any number of languages and in any number of representations, and names that can be used to identify the correct order(s) for the topic within alphabetically sorted lists of topic names applicable to a specific language or domain.

Note: Topic names provide a set of "node labels" for a set of logically related resources which may not have been labelled.

An association links two or more topics. Each link within an association has a role that identifies the reason the topic is part of that association.

Note: Association links provide a set of "edge labels" that describe the relationships between topics (but not between their occurrences).

A facet can be used to assign a property/characteristic to a part of a resource that does not have that property assigned internally. For example, it could be used to identify the language of a particular a piece of text. A facet definition can define a number of different facet values, each of which can be associated with one or more resources.

Where a topic map is designed to identify a subset within a set of topic maps, the "themes to be added" element can be used to identify the scope in which the topics defined in this map, or in any linked map, are to be considered to apply.

2. What can XML Schemas to offer?

Each XML Schema is enclosed within a schema element. This element defines which namespaces, conforming to the [XML Namespaces] specification, are to be used for schemas, and any other namespaces that are to be referred to within the schema.

A complex type definition is defined within an XML Schema *xs:complexType* element. All complex type definitions are assigned class names using the *name* attribute.

Each complex type definition can start with one or more of *xs:annotation* elements that explain the role of the element, and provide other relevant information, such as

development history. Within the annotation *xs:documentation* elements are used to contain language-specific versions of each definition.

Embedded elements are identified by the presence of an *xs:element* declaration within the complex type definition. Within the topic map schema all such declarations are made by reference (using the *ref* attribute) to a global element derived from another complex type definition. The maximum and minimum number of occurrences of the element that are permitted within a topic map can be defined using the *minOccurs* and *maxOccurs* attributes. Optionally, each element declaration can be annotated using the *xs:annotation* element used to annotate the complex type definition.

Groups of elements that may be used interchangeably are enclosed within an *xs:choice* element. Groups of elements that must occur in the order specified are defined within an *xs:sequence* element. The maximum and minimum number of repetitions of the choice or sequence that are permitted within a topic map can, where appropriate, be defined using the *minOccurs* and *maxOccurs* attributes.

Each attribute (property/characteristic) assigned to a complex type definition is defined in an *xs:attribute* declaration. Each attribute is assigned a name (defined using the *name* attribute) which is unique within the complex type definition, and a type based on, or derived from, the [XML Schema: Datatypes] specification. It is also assigned one of the following statements relating to the way in which it is used:

- *use="fixed"*: Only the value defined by the value attribute of the declaration is permitted
- *use="required"*: The developer of the topic map must provide a value for this attribute each time an element of this type is used
- *use="optional"*: The developer of a topic map may optionally provide a value for this attribute whenever an element of this type is used.

Optionally, each attribute declaration can be annotated using the *xs:annotation* element used to annotate the complex type definition.

Within a customized topic map the relationship between the locally named element declaration and the imported complex type definition is declared through a

substitutionGroup attribute, whose value must be the name of one of the abstract element declarations defined for use in topic map schema. The name to be assigned to the instance of the class described within the customized topic map is indicated using the name attribute.

3. What does the XML Linking Language provide?

There are two types of XML links: simple links that reference a single resource and extended links that can reference multiple resources, and define traversal rules (arcs) between them. Only extended links are required to define a topic map.

Elements conforming to the [XML Linking Language] (XLink) specification are identified by the presence of one or more attributes that define the linking properties to be associated with the element. The type of XLink element is identified by a *namespacedtype* attribute from this set. Normally the *xlink* namespace is used for this purpose, e.g. *xlink:type="extended"*.

The following attributes are used to identify the components of an XML extended link that are used within our specification:

- *xlink:type*
- *xlink:href*
- *xlink:role*
- *xlink:title*
- *xlink:label*

The *xlink:type* attribute identifies which type of link object is being defined. Options includes *simple*, *extended*, *locator*, *arc*, *resource* and *title*.

The *xlink:href* attribute is used to identify a resource that is to be "located" by an XLink locator component. It is defined in terms of a Uniform Resource Identifier (URI), as defined in IETF RFC 2396, with an optional fragment identifier based on the [XML Pointer](XPointer) specification.

The optional *xlink:role* attribute can be used to provide a pointer to a resource that explains the role played by the link. It is defined in terms of a URI, with an optional fragment identifier based on the XPointer specification.

The optional *xlink:title* attribute can be used to describe the meaning of a link or resource in a human-readable fashion. It should contain a string that describes the resource. An alternative way of defining titles is by use of an *xlink:title* embedded element.

Note: Our specification uses embedded xlink:title elements in preference to attributes of the same name.

The optional *xlink:label* attribute is used to identify locators that play specific roles within a traversal rule (arc). (The way in which labels and traversal rules should be applied within a topic map is not defined in our specification. Developers of topic map schemas based on the complex type definitions defined in our specification may choose to add them if they are deemed appropriate.)

4. XML Complex Types for defining Topic Maps

Each schema used to define a topic map must be stored within an XML Schema *xs:schema* element. This element shall define the namespaces being used within the definition. Users can assign any namespace name they like to elements, but they must declare at least two namespaces, one of which references the XML Schema specification using the URI *http://www.w3.org/2000/10/XMLSchema* and the other of which references the XML Linking Language specification using the URI *http://www.w3.org/1999/xlink*. Users may choose to support the definitions for XML Linking Language attributes, and the XML Language attribute, from an external source.

The complex type definitions used to define reusable data types within a topic map are named using the name of the topic map component followed by the word *Type*. For each complex type definition there is an abstract element based on this definition which forms an architectural base for elements conforming to the type definition. All abstract element definitions use names consisting of *Abstract* followed by the name of the type used to define the element.

A full definition of the schema, together with examples of its application within Diffuse Topic maps can be found at <http://www.diffuse.org/TopicMaps/schema.htm>.

4.1. The Topic Map Complex Type

A topic map consists of a set of topics, associations, facets and added theme elements that are used to manage a set of terms relevant to a particular knowledge domain.

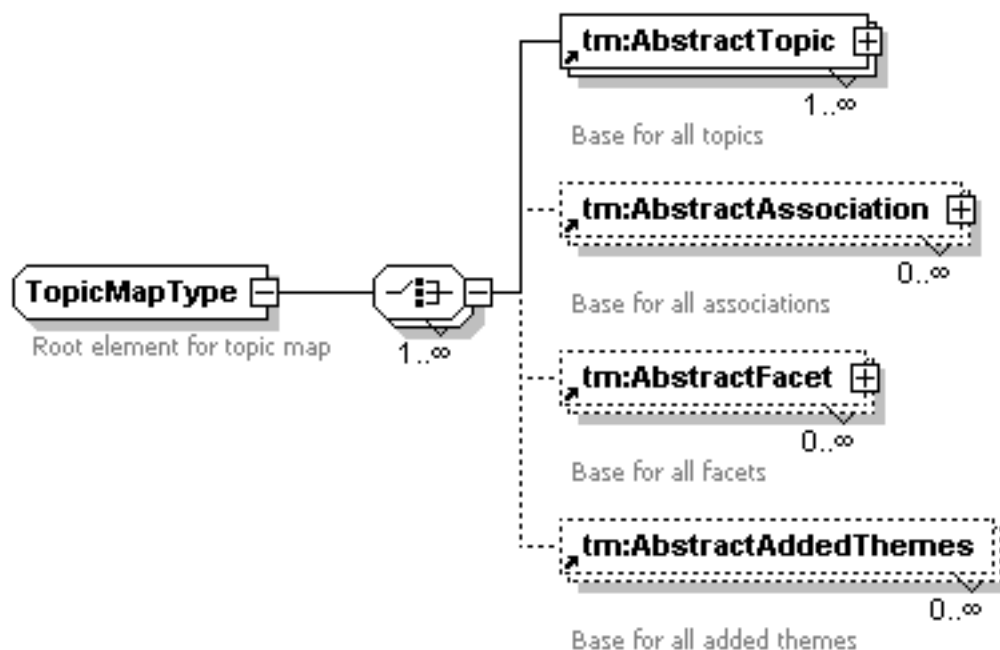


Figure 2. Figure 2

Structure of Topic Maps

Each topic map must start with an element that is derived from the *TopicMapType* complex type definition shown above. This element must contain one or more of the following types of embedded elements:

- *Topic*: At least one topic must be defined within each topic map
- *Association*: Describing the relationship between two or more topics

- *Facet*: Describing a characteristic to be assigned to one or more resources
- *AddedThemes*: Identifying the relationship between topics maps.

The *AddedThemes* attribute associated with the *TopicMapType* complex type definition allows one or more of the topics within the topic map to be used as a general theme for the topic map.

4.1.1. The Topic Complex Type

A Topic assigns a set of topic names to a set of resources that relate in some way to the meaning normally associated with the names.

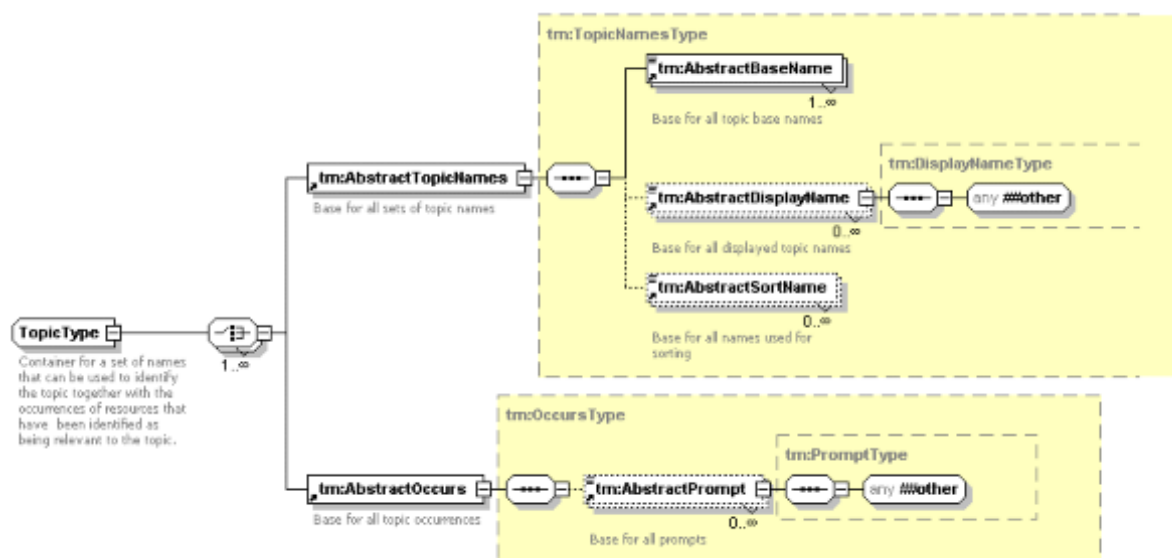


Figure 3. Figure 3

Structure of Topics

A subject descriptor is a reference to a positive, unambiguous, indication of the identify of a subject. For example, it could be a reference to some descriptive text, a Dewey Decimal Code or a Universal Decimal Classification. A public subject descriptor is a subject descriptor which is designed to be used as a common referent of the identity attributes in many topic maps.

The theme of a topic, topic name, occurrence or association is the set of topics

defined in the scope attribute of the element, together with any topics identified as being additional themes for part or all of the topic map (e.g. by an AddedThemes attribute or element).

Each element used to define a topic within a topic map must be derived from the *TopicType* complex type definition shown above. This element conforms to the definition of an extended XML link.

At least one set of topic names or occurrences must be defined for each topic, but there is no restriction on how many sets of names or instances of occurrences may be provided for each topic. Each of the occurrences defined in the topic has each of the names that have the same scope assigned as one of its topic identifiers.

Each topic must be assigned a unique identifier as the value of its *id* attribute so that it can be referenced by associations, etc. This unique identifier must be a valid XML name.

Each topic can optionally be assigned a reference to a subject descriptor as the value of its *identity* attribute. This XML pointer identifies the relevance of the topic by reference to terms that may not be defined as part of a topic map. Any two topics that have the same URI as the value of their *identity* attribute are considered to be equivalent to a single topic that is the union of the contents of the two topics, and of any associations that reference them.

Where a topic is an instance of an existing topic that relationship can optionally be recorded in the *topics* attribute. The value of this attribute must be a list of valid URI references whose fragment identifier identifies the unique identifier assigned to a topic definition.

Note: References to identifiers of topics within the same map must begin with a # to indicate that the entry is a URI rather than a name token.

Where the names and occurrences contained within the topic element are only relevant within a specific knowledge domain, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes of all names and occurrences defined by the topic.

4.1.1.1. Defining Topic Names

A topic name is a string used by a computer or human to distinguish one topic from another. There are three types of topic names: text-only base names, human-friendly display names (which can include images) and computer-readable sort names.

Each element used to define the set of names to be used to identify a topic within a topic map must be derived from the *TopicNamesType* complex type definition.

Each set of topic names must include at least one subelement that conforms to the *BaseNameType* complex type definition (see below). Alternative versions of base names may also be supplied for use in different languages or for use within different scopes (knowledge domains).

Where appropriate, a more human-friendly displayable form of the name may be assigned to the topic for use within a specific language/domain by one or more elements conforming to the *DisplayNameType* complex type definition (see below).

Where the base name is not suitable for correctly ordering the topic within alphabetical listings in one or more of the specified languages/domains, an element conforming to the *SortNameType* complex type definition (see below) can be assigned to the topic.

Where all the entries in a set of topic names are only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes of all names defined for the topic.

Note: If more than one set of topic names is assigned to a topic the names they contain should have different scopes.

Where two topics share the same name within exactly the same set of scopes, the names and occurrences applicable to each scope are treated as a single set of names/occurrences.

4.1.1.2. Base Names

A base name is a string used to distinguish one topic from another.

Each element used to define a base name for the recognition of a topic within a topic map must be derived from the *BaseNameType* complex type definition.

If multiple base names are assigned to a topic each such name should be relevant to a specific language or knowledge domain.

Where a base name is only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes of the base name, in addition to those assigned to any enclosing elements conforming to the *TopicNamesType*, *TopicType* or *TopicMapsType* complex type definitions by their *scope* or *AddedThemes* attributes.

Where a base name is only significant for a specific language the *xml:lang* attribute should be used to identify the language used in conformance with IETF RFC 1736.

4.1.1.3. Display Names

A display name provides a more user-friendly format for the topic for use within a particular language/scope.

Each element used to define displayable name for the recognition of a topic within a topic map must be derived from the *DisplayNameType* complex type definition. If multiple display names are assigned to a topic each such name should be identified as being relevant to a specific language or business domain.

A display name may consist of text or any other element that is not defined within the same namespace as a topic map.

Where a display name is only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes of the name, in addition to those assigned to any enclosing elements conforming to the *TopicNamesType*, *TopicType* or *TopicMapsType* complex type definitions by their *scope* or *AddedThemes* attributes.

Where a display name is only significant for a specific language the *xml:lang* attribute should be used to identify the language used in conformance with IETF RFC 1736.

4.1.1.4. Sort Names

A sort name is used to correctly identify the sequence in which a topic should be listed when this is not directly indicated by the base name. For example, if the base name is Charles VI the sort name might be Charles6 so that similarly named kings will be listed in the correct order, and Charles VI will not appear after Charles IX.

Each element used to define the sorting order for a set of topics within a topic map must be derived from the *SortNameType* complex type definition. If multiple sort names are assigned to a topic each such name should be identified as being relevant to a specific language or business domain.

Where a sort name is only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes of the sort name, in addition to those assigned to any enclosing elements conforming to the *TopicNamesType*, *TopicType* or *TopicMapsType* complex type definitions by their *scope* or *AddedThemes* attributes.

Where a sort name is only significant for a specific language the *xml:lang* attribute should be used to identify the language used in conformance with IETF RFC 1736.

4.1.1.5. Occurrences

An occurrence of a topic identifies which parts of a resource are related to the topic, and the role that part of the resource plays with respect to the topic.

Each element used to identify occurrences of a topic from a topic map within a web resources must be derived from the *OccursType* complex type definition. It can optionally contain one or more elements conforming to the *PromptType* complex type definition.

Each element conforming to the *OccursType* complex type definition is an XLink locator. It must have an XLink *href* attribute that identifies a resource whose content is related in some way to the topic. Where only part of the resource is relevant for the topic a fragment identifier that is conformant with the XPointer specification may be used to identify the relevant part of the resource.

Note: Because XML Links do not allow embedded XML Links to be used as locators of resources, occurrences declared using this schema can only be assigned to a single resource. (ISO 13250 allows an occurrence to identify a set of resources.)

Each occurrence type element may optionally be assigned an XML name that identifies the role played by the occurrence with respect to the topic. If no value is supplied for the *xlink:label* attribute the name assigned to the element is taken to be a sufficient label.

Note: The general-purpose xlink:label has been used in preference to a specialist occurrence role attribute (e.g. occrl in ISO 13250) to ensure maximum possible reuse of in-built XLink functionality. It is equivalent to the role attribute used on higher level elements.

Where there exists a topic that defines a set of names relevant to the role, and/or occurrences explaining it, the *xlink:role* attribute can be used to reference the unique identifier of this topic.

Note: This attribute serves the same purpose as the type attribute in an ISO 13250 occurs type element. It represents a restriction on the xlink:role attribute in that it restricts the type of resource that can be referenced to one that conforms to the TopicType complex type definition.

Where an occurrence is only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes associated with the occurrence, in addition to those assigned to any enclosing elements conforming to the *TopicType* or *TopicMapsType* complex type definitions by their *scope* or *AddedThemes* attributes.

Where the content of the occurrence is in a language other than that used for the bulk of the topic map the *xml:lang* attribute should be used to identify the language used within the resource in conformance with IETF RFC 1736.

4.1.1.6. Prompts

A prompt provides a textual clue to the user as to what he will see if he chooses a

particular resource.

Each element used to prompt the user to select a locator within a topic map must be derived from the *PromptType* complex type definition. It will be considered to be a title element by XLink processors.

Prompts may contain text and any element that does not conform to one of the complex type definitions defined in this specification.

Where the content of the prompt is provided in more than one language the *xml:lang* attribute should be used to identify the language used in conformance with IETF RFC 1736.

4.1.2. The Association Complex Type

An association link expresses relationships between topics. Each such relationship is named (typed).

Note: Association links describe the relationships between topics rather than between specific occurrences of topics within resources.

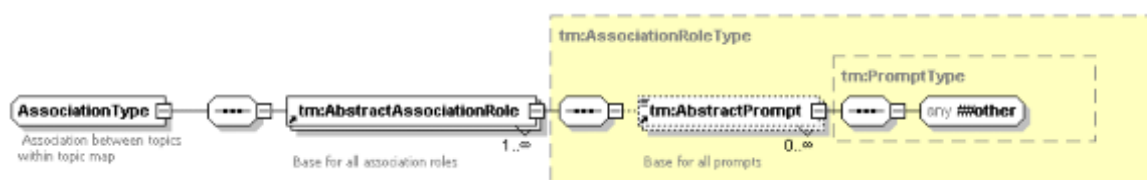


Figure 4. Figure 4

Structure of Associations

Each element used to define a relationship between two or more topics within a topic map must be derived from the *AssociationType* complex type definition. This element conforms to the definition of an extended XML link.

Each association link may optionally be assigned an XML name that identifies the role played by the association. If no value is supplied for the *role* attribute the name assigned to the element is taken to be a sufficient label.

Where an association is an instance of a type of association that has already been named/described in an existing topic, the relevant topic can be identified using the *xlink:role* attribute. The value of this attribute must be a valid URI reference whose fragment identifier identifies the unique identifier assigned to the relevant topic.

Note: The general-purpose xlink:role has been used in preference to a specialist association type attribute (e.g. type in ISO 13250) to ensure maximum possible reuse of in-built XLink functionality. It represents a restriction on the xlink:role attribute in that it restricts the type of resource that can be referenced to one that conforms to the TopicType complex type definition.

Where an association is only relevant within specific knowledge domains, the set of topics that identify relevant domains can optionally be recorded in the *scope* attribute. The referenced topics become one of the themes associated with the association, in addition to those assigned to any enclosing elements conforming to the *TopicMapsType* complex type definition by its *AddedThemes* attribute.

4.1.2.1. Association Roles

An association role identifies the role played by a specific linked resource within a particular association.

Note: An association role can be thought of as an edge label that links two topics to prompt users when they are navigating between topics in a topic map.

Each element used to describe the role of a topic within an association relationship must be derived from the *AssociationRoleType* complex type definition. Each element conforming to the *AssociationRoleType* complex type definition is an XLink locator. It must have an XLink *href* attribute that identifies a topic defined in a topic map, using the topic's unique identifier as the fragment identifier of the URI.

Note: Because XML Links do not allow embedded XML Links to be used as locators of resources, association roles declared using this schema can only be assigned to a single resource. (ISO 13250 allows an association role to identify a set of resources.)

Each association role element may optionally be assigned an XML name that identifies the role played this part of the association. If no value is supplied for the

xlink:label attribute the name assigned to the element is taken to be a sufficient label.

Note: The general-purpose xlink:label attribute has been used in preference to a specialist anchor role attribute (e.g. anchrole in ISO 13250) to ensure maximum possible reuse of in-built XLink functionality. It is equivalent to the role attribute used on higher level elements.

Where there exists a topic that defines a set of names relevant to the role, and/or occurrences explaining it, the *xlink:role* attribute can be used to reference the unique identifier of this topic.

Note: The general-purpose xlink:role attribute has been used in preference to a specialist association type attribute (e.g. type in ISO 13250). It represents a restriction on the xlink:role attribute in that it restricts the type of resource that can be referenced to one that conforms to the TopicType complex type definition.

4.1.3. The Facet Complex Type

A facet link assigns a property/value pair to resources that currently do not exhibit that property.

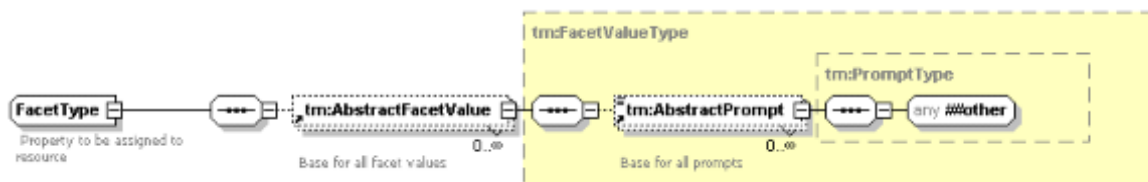


Figure 5. Figure 5

Structure of Facets

Each element used to define a facet (property) to be assigned to a resource must be derived from the *FacetType* complex type definition. This element conforms to the definition of an extended XML link. The element contains a number of elements conforming to the *FacetValueType* complex type definition that assign specific values for the property to a specific resource.

Note: Facets with no currently assigned facet values can be defined as part of the generation of a topic map to indicate that values are expected to be assigned at a

later date (or have been assigned in the past).

Each facet may optionally be assigned an XML name that identifies the name of the property being assigned to resources. If no value is supplied for the *role* attribute the name assigned to the element is taken to be a sufficient label.

Where a facet is an instance of a topic that already exists as part of a topic map, the relevant topic can be identified using the *xlink:role* attribute. The value of this attribute must be a valid URI reference whose fragment identifier identifies the unique identifier assigned to the relevant topic.

Note: The general-purpose xlink:role attribute has been used in preference to a specialist facet type attribute (e.g. typein ISO 13250) to ensure maximum possible reuse of in-built XLink functionality. It represents a restriction on the xlink:role attribute in that it restricts the type of resource that can be referenced to one that conforms to the TopicType complex type definition.

4.1.3.1. Facet Values

A facet value assigns a specific value of a property to a resource, or to part of a resource.

Each element used to assign a specific value for a facet to one or more resources must be derived from the *FacetValueType* complex type definition. Each element conforming to the *FacetValueType* complex type definition is an XLink locator. It must have an *XLinkhref* attribute that identifies a resource to which the property and value are to be assigned. Where only part of the resource is to be assigned the property a fragment identifier that is conformant with the XPointer specification may be used to identify the relevant part of the resource.

Note: Because XML Links do not allow embedded XML Links to be used as locators of resources, facet values declared using this schema can only be assigned to a single resource. (ISO 13250 allows a facet value to be assigned to a set of resources.)

The value to be assigned to the property identified by the enclosing facet type element can be specified as in the *value* attribute. If no value is specified the name of

the element is taken to be the value to be assigned to the property.

Where there exists a topic that defines a set of names relevant to the value, and/or occurrences explaining it, the *xlink:role* attribute can be used to reference the unique identifier of this topic.

Note: The general-purpose xlink:role attribute has been used in preference to a specialist facet value type attribute (e.g. type in ISO 13250) to ensure maximum possible reuse of in-built XLink functionality. It represents a restriction on the xlink:role attribute in that it restricts the type of resource that can be referenced to one that conforms to the TopicType complex type definition.

4.1.4. Added Themes

An added theme is a topic that is to be added to the list of scopes applied to a particular class of elements throughout one or more topic maps.

Each element used to indicate topics which define the scope of one or more types of elements in this or another topic map must be derived from the *AddedThemesType* complex type definition. The element is an empty element which has no embedded content.

The *themes* attribute contains a space separated list of URIs whose fragment identifiers identify topics in one or more topic maps that are to be used to assign additional themes to those listed in the *scope* attribute of one or more types of topic map elements.

The *topicmaps* attribute contains a space separated list of URIs of documents that conform to the *TopicMapType* complex type definition to which the identified themes are to be added to all elements. If no value is assigned to this attribute the topics are added to the topic map containing the element conforming to the *AddedThemesType* complex type definition.

The optional *assignto* attribute can be used to identify specific types of topic maps elements to which characteristics are to be added. More than one value can be specified in a space separated list.

4.2. Customizing Topic Maps

A topic map application must define its own XML schema. This schema will need to include the type definitions and element declarations used to define the base complex types. It will also need to assign local names for each of the abstract element declarations using to form substitution groups of local elements that map to the abstract element declarations.

[Figure 6](#) shows the structure of the schema used to manage the topic maps used by the Diffuse project.



Figure 6. Figure 6

Structure of Diffuse Topic Maps

Elements with names beginning with the *tm:* namespace are abstract elements defined using the complex type definitions detailed above. Names with no namespace are the ones used to define Diffuse topic maps. Each such element is derived from the abstract element the name points to in the diagram. Boxes surrounding each set of elements indicate which complex type definitions they are derived from.

Figure 7 shows the start of a typical display for a Diffuse topic map.

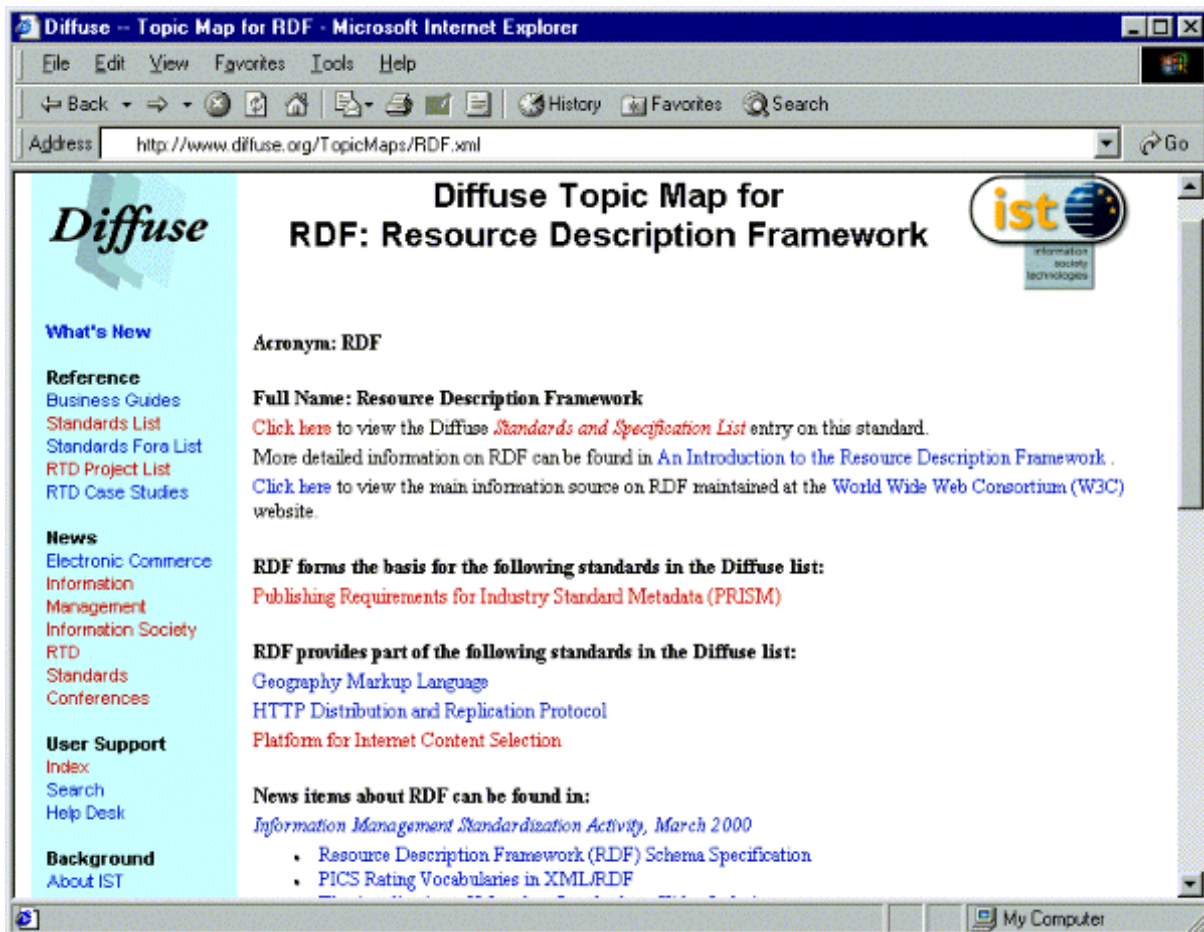


Figure 7. Figure 7

Diffuse Topic Map Display

Biography

Martin **Bryan**

Technical Manager

The Diffuse Project

The SGML Centre

Churchdown

United Kingdom

Email: mtbryan@diffuse.org

Martin Bryan - From a printing and publishing background, Martin Bryan started to work with SGML in 1986, writing the first textbook on the subject in 1987. After a spell as SGML Products Manager for the Sema Group, an international consultancy group, he went on to set up his own company, The SGML Centre, in 1992. Since 1996 Martin has been working with the European Commission to enhance awareness of standards relating to the Information Society, first through the Open Information Initiative, and since February 2000 through the Diffuse Project. Other standardization activities relate to the introduction of XML into the EDI arena, particularly through the ebXML initiative. Martin also chairs the CEN/ISSS EC Workshop working group on Defining and Managing Semantics and Datatypes.