

TMDM

Last chance to change

Status

- **Last vote**
 - Yes, without comments: Canada, Italy, Korea, Netherlands
 - Yes, with comments: Norway, UK, US
 - No, unless comments accepted: Japan (email not received by secretariat)
 - TMDM therefore now approved as FCD
- **Quality of comments received was high**
 - most of the received comments deal with wordsmithing, and so will not be covered in this meeting
 - many will be rejected for technical reasons already discussed in meetings; these will therefore not be covered in the meeting either
- **Comments from Robert Barta¹ are also included**
- **This means that this is the last chance to make non-trivial changes**
- **After this, TMDM will be pretty much cast in stone...**

¹ <http://isotopicmaps.org/pipermail/sc34wg3/2005-April/002609.html>

General: URI vs IRI

- **Japan**
 - "URI should be changed to IRI because an IRI is a sequence of characters from the Universal Character Set (Unicode/ISO 10646) and allows to use non-Latin scripts in it."
- **IRIs are the same as URIs, except that non-Latin characters can be written directly**
 - in URIs they must be UTF-8-encoded and escaped with %xx
- **Effectively, this is an improvement over the current version**

Resolution

- **Editors happy to change the normative reference**
 - However, the IRI identifying the URI/IRI datatype is defined by XML Schema
 - We don't think XML Schema has the right datatype for this...
- **Editors instructed to contact editors of XML Schema part 2 and find out whether they will make a datatype for IRIs**
 - Once such a datatype exists we will use it instead of anyUri



Clause 3.14-3.16: published subject *

- **UK**
 - "Why are these defined? Surely 13250-2 only needs to be concerned with subject identifiers and subject indicators, published or not."
- **This is a carry-over from ISO 13250:2000, which defined "public subject descriptor" in much the same way**
- **The terms are generally useful to distinguish published subject indicators from unpublished subject indicators**

Resolution

- **The editors agree to remove 3.14-3.16 from part 2**
- **Section 7.1 will be rewritten to refer only to "subject identifiers"**
- **A note referencing the OASIS PubSubj TC recommendation will be added**

Clause 4.4: Datatypes

- **Norway**
 - "It should be stated explicitly that the datatype URIs are considered to be PSIs for the respective datatypes, not just 'identifiers'."
- **UK**
 - "[datatype] property. Value should be described as a string identifier."
- **Barta**
 - "The datatype is a string. And then it is a locator one sentence later. I guess it is not a topic for a good reason?"

Resolution

- **The UK's comment is rejected**
 - IRIs are preferred as they help ensure that datatype identifiers are globally unique
 - they are also preferred in order to remain compatible with the XML and RDF families of standards
 - if they are IRIs datatypes can also be reified
- **Norway's comment is accepted**
 - this allows datatypes to be reified
- **The editors will add prose making it clear that the datatype IRI is to be considered a subject identifier**
 - however, the connection remains a loose one in the model
 - merges of reified datatypes will have no effect whatsoever on processing

Clause 4.4: Datatypes

- **US**
 - "Doesn't defining string and using the W3C definition give us two definitions of string? BTW, the XML infoset defines 'null' (...) so wouldn't it be cleaner to simply add 'null' to the list and avoid defining string/null, etc.?"

Resolution

- **Editors will investigate whether xsd:string and TMDM:string are equivalent**
 - if they are, the text can stay as it is
 - if they are not, corrective action will be taken
- **Editors will add text to make it clear that the XML Infoset's "No Value" is the same as the TMDM's "null"**

Clause 5.1: Source locators

- **"The UK would like the editors to review whether clause 5.1, Source locators, should be an integral part of the data model."**
- **Source locators are currently used**
 - when referencing topics by id in queries, schemas, etc
 - to define reification
 - to merge topics during deserialization
 - to support "late merge"
- **However, they are not without their problems**
 - usually file URIs which cannot be guaranteed to be unique in P2P situations
 - don't feel quite right
 - bloat topic map instances, and slow down imports considerably
- **So, what do we do?**

Resolution

- **The editors will change the term "source locator" to "item identifier"**
 - description of use cases for the construct should be added to the definition to make the use and semantics of this construct clearer
 - the property name will change (obviously)

Related issue: Reification

- **Graham**
 - reification is too fundamental to be a computed property
 - having the computation effectively be a lookup is even worse
 - this is only a hangover from XTM 1.0
- **Making reification a fundamental property also means**
 - item identifiers can be freely stripped out of topic maps by applications which do not want them
 - subject identifiers assigned only to achieve reification can be removed during deserialization
- **Proposal**
 - [reified] and [reifier] are no longer computed (contain items as before)
 - on XTM deserialization they are populated directly; <subjectIndicatorRef>s which are used for reification are discarded
 - change definition of [reified] to say "the topic map construct that is reified by this topic"

Clause 5.2: The topic map item

- **Barta**

- "Base locator: I would assume - from the name - that it is an absolute URI which allows all relative URIs in the map to be made absolute.

Why is it there in the model and what reason could there be that this is NOT done at deserialisation (or construction) time?

The only reason can be that the base locator is modified in a TM instance changing all relative URIs in there. I cannot see any use case for this.

It can be NULL. So what does that mean for relative URIs? Is such a map meaningful?"

- **Norway**

- "The semantics of the base locator property need to be defined more precisely."

Resolution

- Goodbye, [base locator]

Clauses 5.4, 5.6, 5.7, and 5.8

- **Barta**

- "In a couple of places the TMDM allows to have NULL as value, even if there were a choice for a reasonable default. So, for instance, if an occurrence does not have a type, it is an 'OCCURRENCE' or if a topic is not an instance of something, then lets make it a 'THING' or 'TOPIC'.

Everyone who defines a mapping from TMDM into something else needs a base ontology for the TMDMish things (Lars needs it for his Q, I need it for \Tau). It is probably not good if we all use different ones.

If that is done clever, we could reuse it for TMQL as well."

- **Norway**

- "The [type] property should be required to have a value. Define PSIs for use when deserializing untyped topic names from XTM."

Resolution

- **Change the [type] property to no longer be allowed to contain null**
- **Define subject identifiers for four untyping topics**
 - `http://.../untyped-topic-name`
 - `http://.../untyped-occurrence`
 - `http://.../untyped-association`
 - `http://.../untyped-association-role`
- **Modify XTM deserialization to automatically add these topics**

Clause 7: Published subjects

- **Norway**
 - "The domain topicmaps.org should not be used for PSIs since it is not controlled by SC34. Use PSIs whose form is <http://psi.topicmaps.com/iso13250/foo>."
- **Resolution: accepted**

Clause 7.4: Variant name scopes

- **Norway**
 - "The display name PSI should be removed. It no longer serves any purpose now that the topic naming constraint has been removed."
- **Resolution: accepted**

Clause 7.5: Topic characteristic types

- **Norway**
 - "The unique-characteristic PSI should be removed from TMDM since it belongs more appropriately in TMCL."
- **Resolution**
 - accepted by the working group meeting
 - editors instructed to invite comments from Steve Newcomb to verify that he also accepts this

Clause 7.6: Topic map constructs

- **UK**
 - "Explain why the identified for topic map constructs are there, or else remove them."
- **Norway**
 - "The purpose served by the PSIs defined in this section should be made clear, or else they should be removed from the standard. If retained, they should be defined more carefully."

Resolution

- **Accepted: we remove clause 7.6**

Clause 7.6: Topic map constructs

- **Norway**
 - "Each term in the glossary should have a PSI."
- **Resolution**
 - a new subclause of clause 7 will be added defining these PSIs, with minimal guidance on their use in actual topic maps

New issue: parameter vs scope

- **Why do variant items have a [scope] property when the <variant> element has a <parameters> sub-element?**
 - the <parameters> element specifies the context in which the variant name is appropriate, and this is the same function that scope on topic name performs
 - the semantics of scope on topic name and variant name are clearly the same, especially as variant names inherit the scope of the topic name

New issue: definition of base name

- **We remove the definition of "base name" as a formal term, but maintain prose describing the [value] of topic names as the base form of the name**

New issue: prominence of type-instance

- **Standard should make it clearer that type-instance etc are part of the model**
- **Alternatives**
 - add a NOTE in the definition of association items (5.7)
 - move 7.2 and 7.3 in at the end of clause 5
 - switch order of clauses 6 and 7
 - add 5.9 about these PSIs, and keep 7.2 & 7.3 where they are
- **Proposal**
 - move 7.2 and 7.3 in at the end of clause 5 and add instance diagrams

Resolution

- **Since part 5 is going to have TMDM-in-\Tau+ in an annex, part 2 no longer needs the annex with the TMRM disclosure of the TMDM; it should therefore be removed**