

# Vocabulary Maintenance Specification (VOCAB) Task Group meeting agenda (Google Hangout)

Meeting time: 2016-05-04 21:00 UTC

<https://hangouts.google.com/call/rm6yosa3afgttcnzdx373tuoqe>

Prepared by Steve Baskauf, TG convener

Participating in the call: Steve Baskauf, Stan Blum, Bob Morris, Joel Sachs, Greg Whitbread, John Wiczorek

Annotations highlighted in yellow have been inserted to include notes about the discussion that took place during the call.

## Purpose

1. Determine whether the draft Standards Documentation Specification at <https://github.com/tdwg/vocab/blob/master/documentation-specification.md> is sufficient to edit and move forward, or whether we need to go back to the drawing board.
2. Close as many open issues as possible in the time allowed.
3. Discuss briefly the way forward for the Vocabulary Maintenance Specification.

## Plan of attack

1. *Thumbs up or down on the draft.* If thumbs up, proceed to next agenda items to enable editing the draft with the goal of submission ASAP. If thumbs down, then spend most of the hangout discussing what needs to be done to achieve forward motion on a new draft.
2. *Close without objection.* Determine whether there are any objections to the recommendations for closing Issues # 19, 24, 25, 28, 29, 30, 33, and 41. Please make up your mind before the meeting so we don't have to spend more than a few minutes on this. (See the section below for details.)
3. *Close with brief discussion.* Discuss briefly (< 5 minutes) how to close issues #31, 32, and 34. Again, come to the meeting with an opinion already formed and we won't have to spend much time on these. (See the section below for details.)
4. *Discuss major issue.* Attempt to achieve clarity on the issue of what parts of standards are normative and non-normative, and how changes to those parts would be related to the as yet undefined change policy of the Vocab Maintenance spec (but using the DwC Namespace policy as a template). What kinds of changes trigger version changes for terms and documents, and what version changes trigger a new release of the whole

vocabulary or standard? (See the section below for details.) Discuss until clear or 21:50, whichever comes first. If 21:50 comes first, then decide whether to schedule another meeting, go overtime, or discuss via email.

5. *Discuss the way forward for the Vocab Maintenance spec.* I'm stuck on this one because I don't know what consensus (if any) was achieved at the last TDWG meeting about the roles of Task Groups, Interest Groups, the TAG, and the Executive in maintaining vocabularies. The current DwC Namespace Policy has a major role specified for the TAG. Should that stay, or instead be placed on the Interest Group maintaining the standard? Or the Process IG? Currently, John Wieczorek as the DwC Task Group incarnate plays the role of the TAG in the DwC Namespace Policy. But according to the TDWG Process, Task Groups should go away when their task is completed and their chartering Interest Group should shoulder the responsibility of maintaining the standard. This is not what has been happening; is it going to be changed or fixed? Possible outcomes are to achieve clarity in a very sudden way, schedule another meeting focused specifically on this issue, or something else I'm not thinking about.
6. (suggested by Bob) *Maybe discuss issue raised in the spec: "[How do we decide the level at which contributors are acknowledged? For something like best-practice guides, it's clear. But what about vocabularies and term lists? [We didn't get to discussing this; it (Issue #37) still needs to be resolved.] Also, does TDWG have a policy that vocabularies (as opposed to descriptive documents) should be licensed CC0? If not, it should!]" just before Sec 4.2.1 [The issue of legal notices and CC0 licensing for vocabularies was discussed and Stan is going to look into it (more below).]*

## Sequential list of Issues to be discussed

Without objection, close the following issues according to the recommendations given:

**Close Issue #19** by including a subsection of section 3.2 indicating that the all-caps words of RFC 2119 should be used only in cases where deviation from the recommendations would break applications, prevent interoperability, or something to that effect. They should not be used in circumstances where best practices are evolved by a community or where there is acceptable variation in use. **[Leave open]** Joel Sachs will look into this further and enter comments in the issue tracker.

**Close Issue #24** by adding text to section 3.2.3.4 (the footer) indicating that a preferred citation in human-readable form and in one or more standardized citation export formats (but not say what they are - this could change over time). I think it's fine not to specify since this would

probably be managed as part of the repo/document management process anyway. [Comment from Stan inserted here: "This suggests that a separate document could/should be created to represent (provide guidance about) current best practice for documenting a standard. Agree?". Response from Steve: Yes, I think a separate document would be fine. The main way that the GitHub maintenance system would interface with this is with respect to what we call a "version" of the standard. "Version" is the generic term used in the spec, but I've been assuming that on the whole standard level, a release would be equivalent to a "version".]

**Close Issue #25** with the understanding that no specific format be specified for human-readable documents. Edit 3.3 and maybe 2.1.2 and 2.2.4 to reflect this.

**Close Issue #28** by going with Bob's suggestion of following the W3C terminology. DCMI is similar except for using "Replaces" instead of "Previous version". In RDF use the properties as already described in the draft.

**Close Issue #29** by removing any specification about name format. We should just have the person list their full name as they typically use it professionally.

**Close Issue #30** Get rid of any reference to hyperlinking to sections from the TOC; leave this up to the navigation system of the delivery system.

**Close Issue #33** Be silent about URI construction and let the decision be made at the time of new term addition based on what makes sense at that time.

**Close Issue #39** (Default serialization). With respect to the specific issue of default serialization, I think it's best to be silent about that and remove text about it. We specify what triples should be included in the machine-readable metadata, but the serialization is an implementation issue that we don't get into.

**Close Issue #41** by going with John's suggestion that the dcterms:modified date be used to indicate the date of death of a term that was superseded or deprecated, with that date corresponding to the date of the decision that killed it. This may be a problem for some older terms, but it could certainly be followed for all term changes in the future. It becomes a technical issue to look at the former terms and figure out what problems (if any) exist with the previous use of dcterms:modified.

## Hopefully close after brief discussion:

**Discuss briefly Issue #31** about motivation/rationale/justification and reach some consensus. Do we care about delineating very specific section names or can we just tell authors they need to include the purpose of the standard in the introduction? We decided that the less prescriptive approach was fine: the introduction should at a minimum describe the purpose.

**Discuss briefly Issue #32** (TDWG legal statement doc) and #38 (Licensing for vocabularies): Each document should have a copyright statement and licensing statement which would correspond to what we put in the RDF metadata. Beyond that, I would like to farm this out to some other part of TDWG to get back to us and tell us what to do. Similarly, it seems like a CC0 license for vocabularies is the way to go, but I think that's a decision that should be made by the Executive, perhaps after reading Jonathan Rees' comments and talking to Hilmar Lapp. **leave open for now** [Who could we farm it out to?] Stan will contact TDWG exec, Jonathan Rees, Hilmar Lapp, and Dataone folks for recommendations about a TDWG strategy on copyright (IP) for TDWG standards.

**Discuss briefly Issue #34** about including the namespace in the human-readable term name. Do we care? If not, flip a coin and either Bob or John is going to have to edit a document to conform. **We concluded that it would be best to include it, since it would help prevent confusion if the term name were used in a context where terms present from other vocabularies having the same local name.**

## Major point of discussion

**This discussion involves aspects of the following issues:**

- #35 Normative parts of term definitions
- #36 Conventions for separating mutable components related to a standard
- #40 What's a version?
- #41 Documenting the date of death of terms
- #43 Term labels

## Background

The [former draft Standards Documentation Specification](#) defined Type 1 (normative) and Type 2 (non-normative) documents that were part of a standard. Type 3 documents were related to standards, but were not actually part of the standard, so they were not governed by any part of the standards process. In that specification, standards weren't to be versioned, but rather to be replaced by a standard with a different name. This draft was silent about machine-readable documents.

The Term Change Policy of the [Darwin Core Namespace Policy document](#) is modeled after the [DCMI Namespace Policy](#). The categories of changes are based on whether the changes are caused by errata, semantic changes of the terms, or addition of new terms. Minor errata or unequivocal errors are corrected without invoking any process. Major equivocal errors require a public comment period, Executive Decision, and version change of the term (or document). Semantic changes (=changes in the meaning?) of terms require public comment, consensus,

and an Executive decision, and trigger a version change. Term additions require justification, public comment, consensus, and an Executive decision and may trigger version changes for documents. This process does not really fit with the Type 1/Type 2 document system of the draft Standards Documentation Specification, and it implied version changes. It wasn't clear about how the change process applied to normative Type 1 and non-normative Type 2 documents or even whether there was any distinction between them.

The current draft of the Standards Documentation Specification does not revolve exclusively around documents. Parts of documents may be declared normative or non-normative, which could in principle go to the level of parts of metadata about individual vocabulary terms (e.g. the definition is normative, but labels and comments aren't). Presumably, we could continue with the distinction between minor/unequivocal errors and equivocal errors, with the former not triggering any official process. But do we make a distinction between changes that involve normative and non-normative parts of documents? Does the same process apply to term definitions, commentary about the vocabularies, examples, best practices, etc. or do different processes apply to different components of the standards documents? Does creating or correcting a translation trigger any process? Are translated documents part of the standard itself or outside of it? Does changing a comment (e.g. example) in term metadata require the same change process as changing the definition of the term?

What happens if we remove the focus on documents entirely? The same information that is in a human-readable document can be expressed in RDF triples. Yet if those triples are loaded into a triple store, the "document" falls apart and the triples are just part of a large graph. Are some triples normative and others not? What categories of triples require different change processes?

### ***Some categories of triples included in metadata about vocabulary terms***

Note: although I've expressed these as RDF triples, most of the same information is present in the human-readable tables in the quick reference guide.

**Normative** (defines the essential characteristics of the term):

```
rdfs:comment "An identifier for the name usage (documented meaning of  
the name according to a source) of the currently valid (zoological)  
or accepted (botanical) taxon."@en ;  
a rdf:Property ;  
rdfs:subPropertyOf dcterms:identifier ;  
[also would include owl:deprecated "true"^^xsd:boolean; ]
```

**Non-normative informational** (helpful for understanding, but does not define):

```
rdfs:label "Accepted Name Usage ID"@en ;  
dcterms:description "Example:  
\"8fa58e08-08de-4ac1-b69c-1235340b7001\""@en ;
```

```
dwcattributes:abcdEquivalence "not in ABCD" ;  
dwcattributes:organizedInClass dwc:Taxon ;
```

**Relational** (relates the term to its term list, versions, and time):

```
dcterms:isPartOf <http://rs.tdwg.org/dwc/terms/> ;  
rdfs:isDefinedBy <http://rs.tdwg.org/dwc/terms/> ;  
dcterms:hasVersion dwc:acceptedNameUsageID-2009-09-21 ;  
dcterms:replaces dwc:acceptedTaxonID ;  
dcterms:created "2009-09-21"^^xsd:date ;  
dcterms:modified "2009-09-21"^^xsd:date ;  
[would also include dwcattribute:decision properties]
```

The triples in the last category (“Relational”) would be assigned or changed automatically and clearly wouldn’t require any change process. The triples in the first category (“Normative”) would clearly require some change process, since they determine the “meaning” of the term. What about the second category (“Non-normative informational”) ?

### **Versions and releases/dcterms:modified**

The Documentation Spec as it is currently written only talks about versions. However, the GitHub repos revolve around releases, which are essentially new versions of the entire standard. What level of changes in the components of the standard trigger a new release? The RDF Guide and all of the dwciri: terms happened but there was no new release, nor even a decision recorded AFAIK. If there is a minor change to term metadata, is the dcterms:modified value updated? What level of change results in a new term version being issued? Does every new term issuance result in a new version of the vocabulary? Does every new version of the vocabulary or a document in the standard trigger a new release?

### ***Some ideas/thoughts:***

Proposed changes in normative properties could trigger the full term change mechanism (public comment, Executive decision), because such changes could potentially break applications. (Note: this is where Issue #19 about RFC 2119 comes in). This would roughly correspond to the DwC Term Change Policy section 3.3 (Semantic changes). Proposed changes in non-normative informational properties could maybe be handled on the Interest Group level, unless the IG thinks that it should be escalated to the full term change level.

What is the role of examples? What about that change that recommended using “pipe” as the separator for lists? It wasn’t in the term definition, but definitely would have affected existing applications.

What about skos:prefLabel in other languages or changes in the English label? Could these be done at will or is a change process required? I think that the following actions should NOT trigger term any change policy:

- creating a skos:prefLabel in a new language.
- creating a new dcat:Distribution that renders in a form that is substantively the same as a previously approved document (e.g. Github Flavored Markup distribution of a PDF or HTML document, or RDF/Turtle serialization of the same triples in an RDF/XML document).

### **Some notes about the conversation during the call:**

There are essentially three separate but related processes going on here:

1. A documentation process (described by the documentation specification), which includes demarcating what is normative and non-normative in standards documents, how version information is recorded, and how versions are connected to each other and to their current resources. It does not stipulate what should be normative or not, nor how versioning should be managed.
2. A vocabulary maintenance process (described by the vocabulary maintenance specification) which includes decision-making about whether changes should be made to vocabularies and terms within them. This specification would presumably trigger varying levels of oversight depending on the extent to which the proposed changes would affect stability and interoperability of the vocabulary.
3. A vocabulary management process (possibly described by some document, but if so, not one that is included as part of a standard) that would include practical aspects of managing documents, endpoints, GitHub repos, etc., and that would involve generating new versions and representations of documents, and releases of standards “bundles” of documents. The changes that take place would be documented by #1 and in some cases triggered by #2, but the management of those changes would be dictated by practicalities, not by prescribed rules.

Maintaining separation among these three processes, would make completing the two standards tractable. The complications involved in any one of these three processes would not necessarily impede description of the other two.

Given that understanding of the situation, these would be the implications for the open issues that were referenced in the beginning of this section:

Issue #35 (normative parts of term definitions): Section 3.2.1 of the current draft documentation specification says that authors of descriptive documents must indicate which parts of the document are normative and which (if any) are not. Vocabulary descriptions are specified as a special category of descriptive documents, so the same thing applies to them.

Machine-readable representations of descriptive documents provide metadata about the documents, but do not include the full content of the document, so issues of normative vs. non-normative content do not apply to them. Machine-readable representations of vocabularies

(in the form of terms and term lists) should include what is essentially the same information as is included in the human-readable representations. Because of the effective identicalness of human- and machine-readable vocabulary representations, whatever designation of normative vs. non-normative that is made in the human-readable representation should also be made in the machine-readable representation. For example, if the human-readable vocabulary term list document states that the definitions are normative but that the comments are not, then the machine-readable description of the term list should include the same statement in an `rdfs:comment` value. I've added section 4.4.2.1 and an example in 4.4.2.2 to clarify this. I have also removed the text from section 3 that declared that normative content is found only in human-readable documents.

Issue #36 (separating mutable components related to a standard): I have retained the requirement that standards documents be written in English. The text of Section 3.2.4 (Language) now reads:

"Standards documents must be written in English. Translations of standards documents are encouraged, but to simplify management of the standard they will be treated as ancillary documents that are not included in the standard." The text of Sections 4.5 and 4.5.2 (regarding multilingual labels) has been left unchanged. If standards documents and labels are translated into many languages (and I hope they are), requiring that all translations be immediately modified with every change to the standard introduces a burdensome requirement on the standards maintenance process - one that probably can't be sustained given the volunteer nature of TDWG. With respect to changes in download and access URLs, as John W. said in his comment on issue #36, changes in those can be accomplished under an errata section of the change policy.

Issue #40 (versioning): The details of this are problems to be handled as part of the vocabulary maintenance and management processes and aren't a concern of the documentation specification. There is now a section 5 (formerly the first of two badly numbered sections 3.3), which describes in a general way the requirements for archiving standards documents without specifying the mechanics of how that archiving will be achieved. I have removed this issue from blocking revision of the draft documentation specification.

Issue #41 (death dates for terms): This was a technical problem related to versioning and deprecation, but assuming that decisions about how to accomplish versioning are outside of the scope of the documentation specification, it isn't directly relevant here and it has been closed.

Issue #43 (term labels): As described in the comments above related to Issue #36, the text currently specifies that English-language labels should be provided and that they should be language-tagged in the RDF. The text about labels in other languages being outside the standard is unchanged from the earlier draft. There did not seem to be any objection to these specifications so I didn't change them. If people have strong feelings about them, it can come up in public comment.



## The way forward for the Vocabulary Maintenance Specification

What consensus (if any) was achieved about the role of Task Groups, Interest Groups, the TAG, and the Executive in maintaining vocabularies? We can't describe a maintenance process if we don't know whose job it is to carry out the process.

In the discussion of this item, a consensus developed that each vocabulary standard should be maintained by a permanent Interest Group whose sole purpose was to maintain the standard. The members of this IG would perform the functions assigned to the TAG by the Change Policy of the DwC Namespace Policy. The TAG's role would be as it was originally conceived: to oversee the interoperability among all TDWG standards.

This organizational structure would essentially codify the way that vocabulary maintenance has evolved to operate, i.e. that those individuals who have a vested interest in maintaining the success and utility of the vocabulary would be the ones charged to be responsible for the key roles in its maintenance. To achieve this would require the chartering of new IGs to maintain Darwin Core and Audubon Core. The Darwin Core Task Group (whose task of developing DwC has been completed) would be dissolved, as would the MRTG (if it hasn't already). Whether or not the IGs that commissioned the TGs would remain would depend on whether there was sufficient interest among their core members to keep them operating.

There may be other plans afoot about organizational structure coming from the Executive. We need to contact Cyndy to find out what those plans are. Steve will start working on a draft of the Vocabulary Maintenance Specification based on the consensus about roles of IGs, modified as necessary based on what the Executive is working on.