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

Meeting time: 2016-07-25 21:00 UTC

Hangout link: https://hangouts.google.com/call/zulz5dj3wjf4pjzcwczddczpr4e (the link on the calendar also works)

Prepared by Steve Baskauf, TG convener
Participating in the call: Steve Baskauf, Jonathan Rees, Greg Whitbread, Joel Sachs, Stan Blum, William Ulate
John Wieczorek was unable to attend, but in an email expressed support for moving the drafts forward.

Purpose

To determine whether there is a consensus of the Task Group to recommend that the draft Standards Documentation Specification and Vocabulary Maintenance Specification be submitted to the TDWG Executive Committee for appointment of a Review Manager and initiation of expert review.

Standards Documentation Specification draft under consideration:
https://github.com/tdwg/vocab/blob/master/documentation-specification.md

Vocabulary Maintenance Specification draft under consideration:
https://github.com/tdwg/vocab/blob/master/maintenance-specification.md

Possible courses of action:

1. Recommend the drafts as-is.
2. Recommend advancing the drafts after minor revisions (requiring additional discussion via the Issues Tracker, but no additional Hangouts)
3. Not recommend advancing the drafts and carrying out major revisions (requiring another Hangout to consider the revised draft)

About contributors:

Primary authors listed first, followed by other contributors in alphabetical order. “Other contributors” participated in the discussion or suggested text for the document. For the Maintenance spec, I’m not completely clear about the contributions to the existing DwC Namespace Policy document and probably should add the names of all participants in the Hangouts.

Summary of main features of the specifications:

Both specifications:

1. Base the format and processes on existing TDWG precedents where they are working.
2. Change or add processes only where deficiencies have been identified.

Standards Documentation Specification:

1. Human-readable guidelines (Section 3) are based on the previous draft specification written by Roger Hyam that has been serving as the de facto standard. Guidelines for machine-readable metadata have been created (Section 4).
2. It eliminates the focus on specific documents that are designated as Type I (normative), Type 2 (non-normative), and Type 3 (outside the standard). Standards components are now abstract entities (Section 2.1) that can be represented by documents in various human- and machine-readable formats. A single component can be composed of both normative and non-normative parts that are clearly designated (Section 3.2.1).
3. It describes the metadata elements used to describe each kind of standards component (human-readable descriptive documents, vocabularies, term lists, terms, and distributions) in both human- and machine-readable formats. The recommended elements were based on perceived current best practices.
4. It links the standards components according to a hierarchy model using [illegible] (Section 2.2). A machine should be able to discover all of the components of a standard and “know” their type and positions in the hierarchy.
5. It describes a version model similar to what has been implemented in Darwin Core and by the World Wide Web Consortium (W3C) for its documents (Section 2.3). Each resource exists as a “current resource” that persists indefinitely and one or more versions of the resource that may be replaced or deprecated. The current resource and versions are linked using appropriate Dublin Core properties.
6. It describes how controlled vocabularies should be structured and described (Section 4.5.4). There are no current controlled vocabulary standards, although it’s listed as one of the TDWG Standard Categories: Data Standard
7. It assumes a layered approach to vocabulary development and describes how basic term lists can be combined with extension term lists to create ontologies (Section 4.4.2.2)

8. It dictates that a draft of a document created by a disbanded Task Group will be published as Task Group Note with a notation that it is no longer under development. This is similar to the publication of Working Group Notes by the W3C. (Section 5.1)

Vocabulary Maintenance Specification:

1. The responsibility for vocabulary maintenance has been clarified by specifying that a dedicated Interest Group (IG) must be formed with a single, specific charge to maintain the vocabulary (Section 2). In the past the roles of Interest Groups, Task Groups, and the Technical Architecture Group (TAG) have been unclear. The chartering and operation of the maintaining IG falls under the normal TDWG Process. The maintaining IG manages the change process and must review proposed changes annually to prevent them from staying in limbo for long periods of time.
2. The general procedure for changes (Section 3) is modeled after the Change Policy of the Darwin Core Namespace Policy document.
3. A variation of the general procedure is described for making changes to documents associated with a vocabulary or adding new documents to a vocabulary standard (Section 3.4). This has previously been a gray area with no clear guidelines.
4. The specification also assumes a layered approach and creates a requirement that proposals for vocabulary enhancements (coordinated changes that go beyond the addition of individual terms) must document clear, community derived goals (Section 4.3.1) and report implementation experience (Section 4.3.2) before they are formally proposed as additions to the standard and subjected to public comment. There is no existing requirement for this in TDWG, but this requirement is similar to the requirements of the W3C and Internet Engineering Task Force (IETF).
5. The specification lays out three general considerations for changes to a vocabulary standard: demand, efficacy, and stability (Section 3.3.1). In cases where the specification does not provide procedural details, the maintaining TG has discretion to decide a course of action that is guided by these general consideration. I think this is an important feature of the specification, given that past proposals have languished for long periods of time because it was not clear who had the responsibility to make decisions in cases where an exact procedure was not specified. The public comment period and final Executive Committee decision provides checks on this increased power of the Task Group.

Discussion:

1. (Raised by Jonathan Rees in https://github.com/tdwg/vocab/issues/16):
>2.1 'a vocabulary maintenance Interest Group cannot be disbanded
>unless the vocabulary it maintains is deprecated' seems draconian.
>IEEE has lots of orphan specs and people use them (e.g. RFC 2119).
>They are what they are and maybe they don't need active maintenance.

>If they need a revision, a new group of authors can be convened to
>make a new spec.
Maybe we should talk about this one. The problem that TDWG has had in the past was half-finished or unmaintained work, e.g. the TDWG Ontology, the Standards Documentation Specification, NCD draft standard. My thinking was that if a vocabulary was a "living" thing, then somebody should tend it and if a vocabulary (or standard in general) was dead, then people should be told that by a deprecation notification. I wasn't really thinking about the case described here. Perhaps there needs to be some status between "actively maintained" and "deprecated" (i.e. not recommended for use).
Didn't really discuss.

2. (Raised by Jonathan Rees in https://github.com/tdwg/vocab/issues/16):
>3.3.3 'the decision should be reported' seems odd in conjunction with
>'no decision will be recorded in the decision'. Maybe that's what you
>mean, in which case leave it, but it seems weird to have a decision
>history that does not record a decision.
In the third case, the proposal is not resolved and is sent back to the public review stage. Ultimately, there would be either acceptance or final rejection that would be recorded. I guess it wouldn't hurt to record a decision in this third case. I think that decision http://rs.tdwg.org/dwc/terms/history/decisions/#Decision-2011-10-16_3 is something like what Jonathan describes. We could talk about this one.

[JAR: My comment was just a reaction to how the document reads and is more a call for smoother prose rather than a challenge to what's proposed. Perhaps some simple change will help a reader, such as replacing 'No decision will be recorded' with 'The decision will not be recorded'.]

3. (Raised by Jonathan Rees in https://github.com/tdwg/vocab/issues/16):
>3.3.4.1 So individual terms are versioned, not vocabularies, right?
>Does that mean each term is its own Standards Document (with respect
>to the process)?
>I think maybe you mean that term documentation is part of some
>vocabulary specification, and that the vocab spec has to be modified
>according to DOC-SPEC? But which level is invoked at each point in
>the process is pretty muddy to me (on a first reading).
Based on the precedent of DwC, individual terms are versioned. They aren't individual documents, but part of term lists according to the hierarchy model. Vocabularies can also be versioned in accordance with the general version model in the specification. However, the conditions under which changing to a new version of a term would trigger the change in the version of the containing the term are not clear to me. We talked about this in Issue 40, but my take on that was that it was too restrictive on implementers for the specification to dictate how the versioning of vocabularies was managed. It's possible that for efficiency's sake, new releases/versions of the vocabulary might not be released if it was felt that additional changes were likely to be approved soon. If this is too unclear to move forward, then perhaps a re-write of the section would be in order before submission. However, I'm not 100% sure I'd know how to write it at this point.

[JAR: again, my interest here is in how easy the document is to read, which is a second order concern. I have no reason to think there's a problem in what's proposed. I searched the vocab maint document and the word 'hierarchy' does not occur. Maybe a simple reference to the hierarchy model somewhere in 3.3 would do the trick - although I haven't read about the model, so not sure.]

Comment made in the chat window by Jonathon Rees:
People writing clients and want the server to do all the data cleaning, vs. people writing servers and want the client to do all the data cleaning
you're not going to stop them from being unspecific...
tons of people still say "CC-licensed"

Some notes made by memory after the call:

There was some concern expressed about the complexity of the hierarchy model. It was noted that there is the option to create vocabularies under separate standards rather than adding terms to existing ones. That could reduce complexity. It may also be desirable for members of the TAG to help task groups shape their developing vocabularies into conformity with the documentation specification. It is also not necessary for vocabularies to include multiple resources in the layers of the hierarchy model. It is possible to have a standard with one vocabulary composed of one term list. The structure would be considerably less complicated than that shown in the hierarchy model diagram.

It was suggested that we get input from stakeholders who have implementation experience (GBIF, ALA) and from the Infrastructure group who are managing the Github site. However, the standards are written to be somewhat implementation agnostic, so that feedback is probably not imperative ahead of submission for review. Perhaps the expert reviewers can come from those communities.

The consensus at the end of the call was to recommend advancing the drafts to the review manager stage. Before doing so, there will be one more round of comments/edits via either the issues tracker or a Google Doc by the end of the week (Friday, July 29). Steve will revise as necessary and put out a last clean version for the TG to review during the week of Aug 1-5, then forward the request to the Executive.