

Date: 13th November 2016

Summary of the expert review evaluation

This document provides a summary of the (three) expert review reports and is submitted for assessment by the TDWG Executive Committee.

The TDWG Vocabulary Maintenance Specification Task Group (VOCAB)¹ was established in May 2015 and submitted on 2nd August 2016 two standards for review²: (1) "**Standards Documentation Specification**"³, and (2) "**Vocabulary Maintenance Specification**"⁴. After initial approval by the TDWG Executive Committee, the proposed standard documents were submitted for expert review. The review manager was appointed in September 2016.

Three invited expert reviewers reviewed both standards together as one package. The expert peer review was organized as a single-blind review with the names of the expert reviewers undisclosed to the authors. The experts were selected with the aim to maximize diverse and complementary backgrounds and experiences. One of the experts was selected from the agrobiodiversity community, another expert was selected from the Dublin Core community, and the third expert was selected from the Natural History Museum community. Two experts are located in Europe and one expert is based in the United States of America.

The first invitation to the expert reviewers was sent on 29th September 2016 and the final referee confirmation was received on 18th October 2016. The peer review time period was set to one month with deadline 31. October. Some of the three expert reviewers asked for and was granted a one-week extension justified by a particular busy travel schedule during the review period. The first review report was submitted on 7th November 2016 and recommends **accepting conditional revisions**. The second review report was submitted on 11th November 2016 and recommends **accepting with minor revisions**. The third review report was submitted on 13th November 2016 and recommends **accepting** the submission.

My recommendation, based on the reports from the expert reviewers, is to accept the submission while recommending that the authors assess and respond to the recommended revisions suggested by the expert reviewers. The documents are indeed very well made, but also complex and difficult to read and fully understand without deep knowledge of respective technologies. The target audience is authors and curators of TDWG standards, and the prerequisite deep technical understanding might be warranted. However, the authors should aim at readability and at shortening the text.

I praise the authors for the high quality of the standard specification documents!

¹ <https://github.com/tdwg/vocab/blob/master/vmg-charter.pdf>

² <https://github.com/tdwg/vocab/releases/tag/2016-08-02-review-submission>

³ <https://github.com/tdwg/vocab/blob/master/documentation-specification.md>

⁴ <https://github.com/tdwg/vocab/blob/master/maintenance-specification.md>

Expert reviewer 1

Review of the specification «TDWG Vocabulary Maintenance Specification» (EXPERT REVIEWER 1)

This document describes how the TDWG standards must be produced. Two types of documents to describe standards are described in the document: the human-readable documents and the machine-readable documents. I would advise the authors given the complexity of the document, to simplify their examples as well as trying to be more concise in their explanations. The section 2 and 3 are clear, whereas the section 4 is hard to understand.

Expert reviewer evaluation: Accepted, conditional revision

Recommended revisions:

In order to ease the understanding of the different examples, I would recommend to keep only the statements that are directly relevant to what you are looking to illustrate. For instance, example section 4.5.4.1, you don't need to show the version information as this point has been extensively described previously.

In the header section of the standard, I would suggest to add the status of the document in order to be compliant with your own guidelines (section 3.1.5) regarding the human readable documents.

In the specification, you used several times the term "dataset". I believe that you consider a dataset as being a "collections of structured metadata". This term can be added to the section 1.5 "Definitions".

Regarding the use of dcat:Dataset in several of your examples, I do not think that this is semantically correct to type a resource as being both a dcat:Dataset and a skos:ConceptScheme or an owl:Ontology (e.g. examples 4.5.4.1, 4.4.2.3, 4.4.1.1)

Section 4.2.2, I found confusing that you recommended the use of both dc:contributor and dct:contributor⁵. As mentioned on the DC website (<http://dublincore.org/documents/dces/>)

⁵ **Review manger comments:** dc:contributor is used in the table while dcterms:contributor is used in the text below the table. Both dc:contributor and dcterms:contributor is correctly used in the examples in section 4.2.3. See also: http://wiki.dublincore.org/index.php/FAQ/DC_and_DCTERMS_Namespaces

in order to not affect the conformance of existing implementations of "simple Dublin Core" in RDF, the domains and ranges of the dc properties have been added to the dct namespace instead of been directly added to the dc namespace. Moreover, these dct properties are defined as subproperties of the dc properties. Consequently, it is not semantically incorrect to use the dc:contributor with a literal or with an Agent. Having both dct:contributor and dc:contributor sound to me like a duplication of the information.

I also wonder why you decided to not include dct:contributor to the table.

Finally, you recommend to use dc:creator with a literal but you are not mentioning the dct:creator property.

Section 4.3.1, in the example, I do not understand why the resource `<http://rs.tdwg.org/dwc/terms/guides/text>` does not have a owl:versionInfo tag. It is though mentioned in the introduction of the session that *"The property owl:versionInfo, which is expected to have a literal value, SHOULD be used to provide information about the version of a resource."*

Section 4.4, the example is not correct. The `<>` should not be used here:

```
:term dcterm:isPartOf :termList. :termList dcterm:isPartOf :vocabulary.
```

I would also suggest to define the base.

In the example section 4.4.1.1, you are looking to illustrate how vocabularies and term lists must be treated. I understand that vocabularies should not use a trailing slash, contrary to what must be done for term lists. Consequently, I believe that you forgot a trailing slash at the end of the following resource in your example:

```
<http://rs.tdwg.org/dwc/dcmi-terms>
```

Section 4.4.2, I would recommend to get rid of the following paragraph: *"When a preferred namespace prefix is chosen, care should be taken to ensure that there is no collision with other commonly used prefix abbreviations. The namespace IRI will often be the same as the IRI that identifies the term list, although this is not always the case, particularly when hash URIs are used. For example, the SKOS ontology is identified by `http://www.w3.org/2004/02/skos/core`, but the SKOS namespace (usually abbreviated skos:) is `http://www.w3.org/2004/02/skos/core#`."*⁶

Indeed, a prefix can vary in different documents. The machine will still be able to correctly deal with that as it will look for its definition (association prefix-namespace).

⁶ **Review manager comments:** Reviewer#1 recommends in section 4.4.2 to delete the paragraph discussing preferred namespace prefix. An alternative might be to clarify in section 4.4.2 that this recommendation is about providing guidelines when a `vann:preferredNamespacePrefix` is declared. There is a catalog of declared preferred namespace prefixes at: <http://prefix.cc/popular/all.rdf> that might be a useful reference?

About your comment on SKOS, the two URIs are different so they don't represent the same thing. The URI with the trailing # is just a convenient way to access the different SKOS resources. Interesting related reading: <https://www.w3.org/wiki/HashVsSlash>

Section 4.4.2.3, in the example showing how basic and extension term lists can be used to create several vocabularies with differing levels of semantics, the import `<http://rs.tdwg.org/ac/enhanced/>`; can be removed as according to the owl reference "importing an ontology into itself is considered a null action".

Same section (4.4.2.3), the following example can be simplified:

```
### Restriction: Metadata Language Required ###
ac:MediaResource a rdfs:Class;
                  rdfs:subClassOf
                    [ a owl:Restriction;
                      owl:minCardinality "1"^^xsd:nonNegativeInteger ;
                      owl:onProperty <http://rs.tdwg.org/ac/terms/metadataLanguage> ].

### Restriction: Metadata Language Not Repeatable ###
ac:MediaResource a rdfs:Class;
                  rdfs:subClassOf
                    [ a owl:Restriction;
                      owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
                      owl:onProperty
```

I would recommend to replace the min and max cardinalities by owl:cardinality 1. The reading of the example will be therefore eased.

Section 4.5.4.1, page 31, regarding the following resource:

```
<http://arctos.database.museum/guid/MVZ:Mamm:115956#specimen> dwc:disposition "in collection".
```

You consider that this example is an example of the use of a controlled vocabulary term to describe a specimen. By using a literal, I would not consider that as using a controlled value. In the definition of the concept `inCollection`, you assigned an URI to this value. The URI should be used instead of the literal value or, if you prefer to keep the literal value, both of them (URI and literal) should be used.

Moreover, I would argue that the `rdf:value` tag in the skos concepts of your example is unnecessary. What is the value of such a tag compared to the label?

A comment about your statement on the unicity of the terms in the TDWG controlled vocabularies. I believe that forcing the terms to be unique is kind of against the goal of a controlled vocabulary. I acknowledge the fact that controlled vocabularies can be used for indexing purposes and thus having unique terms may ease this task. Having said that, you need to keep in mind that you might lose flexibility regarding the search capabilities.

There are also some minor grammatical and spelling errors in the specification:

- Section 3.2.3.1, *this SHOULD be noted in a Status **note***, with a description of the reason
- Section 4.4.2.2, *list to create **an** semantically enhanced ontology*
- Section 4.4.3, rephrase: *Since term lists are typed as dcat:Dataset, they MAY be assigned any of the properties that are appropriate for instances of dcat:Dataset*
- Some differences exist between the current word document and the document hosted on GitHub (e.g. example in section 4.3.1)⁷:
<https://github.com/tdwg/vocab/blob/master/documentation-specification.md>

⁷ **Review manager comments:** For the expert review, I copied the text description to a Word document to allow the expert reviewers to add comments and suggestions using track-changes functionality (adding comments directly in GitHub would not ensure a single-blind review type). I have added some few new lines, but I believe there are no other differences between the Word document and the text in GitHub!

Review of the specification «TDWG Vocabulary Maintenance Specification» (REVIEWER 1)

This document aims at describing how the vocabularies and their associated documents published under the umbrella of TDWG, must be maintained and modified as needed. For that, the authors precisely and clearly define the process of the evolution of the TDWG vocabularies. This is an important topic which needed to be addressed.

This specification is clear and well written.

Expert reviewer evaluation: Accepted, conditional minor revision

Recommended revisions:

In order to ease the understanding of the specification, I would suggest incorporating a figure to illustrate the different steps of the modification process, similarly to what has been done in this document: <https://www.w3.org/2015/Process-20150901/#rec-modify>.

I would also recommend to use RFC 2119 in this specification. You mentioned that it is good practice to do so in the TDWG Standards Documentation Specification.

In section 2.3 and section 2.4.3, you mentioned the “issue tracker”. Although I do understand that you refer to the GitHub issue tracker, this tool has not been introduced before. Is GitHub a requirement for maintaining the TDWG vocabularies? May other versioning systems be used? I would suggest to define the term “issue tracker” before to use it.

There are also some minor grammatical and spelling errors in the specification:

- Section 3.2.2⁸, first paragraph: a change instead of *an change*
- Section 3.2.2⁹, second paragraph: a consensus has been reached instead of *a consensus has reached*

Can you confirm that in section 4.2.3, the reference to the section 4.2 is correct? *The user feedback reports are also used by the maintaining Interest Group to assist in the decision of whether to advance the formal proposal to public comment (Section 4.2), and as an organized source of information to the community during the public comment period.*

⁸ **Review manager comment** : Should be section 3.3.2 (not 3.2.2)

⁹ **Review manager comment** : Should also be section 3.3.2 (not 3.2.2)

Expert reviewer 2

Review of the specification «TDWG Standards Documentation Specification» and «TDWG Vocabulary Maintenance Specification» (REVIEWER 2)

I've done a quick review of the documentation, and it looks quite competently done.

Expert reviewer evaluation: Accepted, revisions recommended

I took a closer look at the Specification and had two strong reactions. First as a specification, it is exceptionally well crafted and carefully written. Clearly the group tasked with its creation has done a fine job.

My second reaction is a bit more complicated, but before I get to the concerns I have, I'd best give you a bit more background.

[Review manager comment: *This following paragraph has been modified to ensure a single-blind review.*]

You might want to look at the Resource Description and Access (RDA) standard (not to be confused with the standard with the same acronym being developed by the Research Data Alliance). RDA is in the midst of some significant changes in its organization, becoming more global and less Anglo-American. The RDA Registry (<http://rdaregistry.info>) is based on Git and GitHub. You might want to take a look at that site, because it represents the thinking and experience of a small group that has (in my opinion) been working far ahead of most vocabulary developments out there. I'm not a coder, but I have learned a bit about how to 'read' it from a point of view of someone who has been developing vocabularies (and re-developing them, as is not unusual) for well over a decade. The RDA Registry data originates in the Open Metadata Registry (<http://metadataregistry.org>). The newest features enable spreadsheet-based import and export for maintenance and to support the addition of an expanding number of RDA translations built to operate as an integral part of the RDA standard. The RDA Toolkit (the usage instructions) is used to keep both tools synchronized, and data to a tool called RIMMF (RDA in many metadata formats) that allows users to create very sophisticated linked data and share it (albeit in a somewhat limited way).

Report from: EXPERT REVIEWER 2

As I read through your specification, I saw that much of what you've done is based to some extent on the DC standard and its documentation. Interestingly enough, the DC standards have no such specification and the maintenance is accomplished by one person who works almost entirely using text files. DC has a relatively small set of terms and simple relationships so, aside from the fact that only one person knows what's inside the black box (always dangerous, to my way of thinking), it works, for now, anyway.

But it's a very old fashioned way of doing vocabulary development and distribution, difficult to extend, and very hard to manage well, even with a good group of collaborators operating under a strict and well-documented specification. Vocabulary development based on the 'document model' is being phased out in most places in favor of more machine based methods that don't require the heavy duty human efforts you folks seem to favor.

That said, for the short term you have a regime that seems to be working for you, but it is not likely to have much of a shelf life. I'm not sure how this exposition fits into your desired review outcomes, but surely you have a better idea of the implications than I do.

Regards, and thanks again for your patience,

[REVIEWER 2]

P.S. Many years ago I remember having some conference discussions with some Darwin Core users, and suggested a way to avoid repeating the entire Taxonomic hierarchy in every record, but I see from your documentation that it's still part of the standard!

Expert reviewer 3

Expert reviewer evaluation: Accept submission

Review of the specification «TDWG Standards Documentation Specification»
and «TDWG Vocabulary Maintenance Specification» (REVIEWER 3)

I confirm that I now have read the standards and I have no serious concerns. I applaud the authors for their effort and

Accept submission

Best wishes

[REVIEWER 3]