TDWG Technical Architecture Group (TAG)
Online meeting
2022-06-27 13:00 UTC

Agenda/Meeting notes:

Participating:
- Steve Baskauf
- Tim Robertson
- John Wieczorek
- James Macklin
- David Shorthouse
- Ian Engelbrecht
- Camila Plata
- Sarah Vincent
- David Fichtmueller
- Visotheary Ung
- Ben Norton
- Camila Plata
- Lyubomir Penev

Regrets:
- Jeff Gerbracht
- Rod Page
- Raïssa Meyer
- Rob Sanderson

Comments and notes taken during the meeting are in red.
NOTE: please make comments as text directly in the document rather than in marginal comments.

There was interest in having a recording of the meeting for those who couldn't be present to view. It would also help me with note taking. So unless someone objects, Steve will record the meeting for private viewing.

I. Set time and date for next meeting. Suggest Monday, 2022-08-29 at 13:00 UTC. People were OK with that.

II. Charter revisions were approved by the Executive Committee on 2022-06-24 and made in https://github.com/tdwg/website/pull/450 .

III. American Library Association (ALA) Metadata Schema Assessment Framework public review and feedback: https://docs.google.com/document/d/1v6FROf5MvXOE1ys0iOczjIYBR0LRquNn2tx1syP3p7Y/edit Interesting 4-level assessment of metadata schemas based on a number of criteria. May be a useful tool for reflecting on our vocabulary standards.

IV. Feedback to the Executive about the Latimer Core draft submission regarding any major deficiencies that might be an impediment to its successful expert and public review.

A. Executive likely to make a decision on advancing the proposal to appointment of a review manager at 2022-06-28 meeting.
B. Sarah Vincent represents Latimer Core in our group and may be able to address any questions.
C. Comment held over from last meeting's notes: One thing to note is Issue 18 https://github.com/tdwg/tag/issues/18, which notes the potential overlap in the concept of a Collection entity in Audubon Core and Latimer Core. Steve is not aware that there has been any further discussion about this.
Comment from Sarah V: See https://github.com/tdwg/cd/issues/369 and https://github.com/tdwg/tag/issues/18
D. Comment from Steve: After reviewing the submission, I think the status of the guidance documentation (https://github.com/tdwg/cd/wiki) needs to be clarified. Section 1 says "The wiki content is non-normative (i.e., not part of the formal standard) …". This is unclear, since content can be non-normative and included in a standards document, and a document can simply be outside the standard. The distinction is a bit subtle – non-normative content is still governed by rules (very lenient) in the SDS and is under the control of the maintaining Maintenance Group. Content outside the standard isn't governed by anything and could be edited at will by anyone. In this case, the issue would be whether this document is part of the formal review of the standard. If it's non-normative, but included in the standard, it would be. If it's outside the standard, it wouldn't be. I think it's critical that this be clarified prior to the submission going to reviewers. Given that there aren't any other documents describing how LtC will operate, it seems like this doc needs to be included in the standard. However, as a wiki, it is pretty mutable. For reference see Section 3.2.1 of the SDS:
https://github.com/tdwg/vocab/blob/master/sds/documentation-specification.md#32-descriptive-documents and Section 3.4.3. of the VMS:
https://github.com/tdwg/vocab/blob/master/vms/maintenance-specification.md#34-changes-to-documents-supporting-the-vocabulary
Comment from Sarah V: The non-norm docs should be reviewed along with the normative - will update Section 1 to remove the 'not part of the formal standard'. With regard to the fact they're in a github wiki:
  ● The wiki is only editable by github accounts/teams with push perms for the repo - at the moment this is mostly limited to the tdwg CD 'team'. We could limit this to named accounts for members of the maintenance group if that would work better? Steve comment: I think this can be worked out with the review manager once named. Changes can happen, but in coordination with the review manager – should make sure there is a stable version for the reviewers.
  ● Edits to the wiki are recorded and you can diff the versions to see changes made, when they were made and by whom they were made, so there's a paper trail of edits, at least.
E. David F. asked about the URIs not dereferencing. Steve clarified that will happen at the end of the process after ratification.
F. Take note of https://github.com/tdwg/tag/issues/33 regarding boolean values. Can we create an ad hoc committee to come up with a recommendation for

this by our next bi-monthly meeting? Tim and Ben (and anyone else interested) to prepare a proposal.

G. Query from Tim R: use of ltc namespace for Event class - why not just use the dwc namespace? There's some discussion in the issue https://github.com/tdwg/cd/issues/362 . Sarah V. commented that they took the approach that if the use was the same, they kept the namespace, but if they narrowed it, they used the ltc namespace. Ben N. asked if there were a way in the documentation to use some kind of mapping (like SKOS narrow match) to indicate the relationship. Steve noted that the typical way to handle this in the W3C would be to use subclass or subproperty to indicate the relationship. For example, formerly dcat:Dataset was a subclass of dcterms:Dataset (although not in v.2, see https://www.w3.org/TR/vocab-dcat-2/#Class:Dataset). But it is questionable whether this kind of semantics is desirable at the "bag of terms" level. See Section 4.4.2.2 of the SDS https://github.com/tdwg/vocab/blob/master/sds/documentation-specification.md#44-vocabularies-term-lists-and-terms for details. Unless this is deemed to be a TDWG-wide issue, we can just let it be handled during the review process and deal with it on the TAG level if questions are raised about it then. Historically, issues involving borrowed terms were handled on a vocabulary by vocabulary basis, e.g. Audubon Core borrowed many terms from outside of TDWG and sometimes recommended that they be used in ways different from the minting vocabularies. (Indicated in the "usage" field of the metadata, a.k.a. skos:scopeNote).

H. No one felt there were any red flags sufficient to delay the appointment of a review manager and Steve will report this at the upcoming Executive meeting (day following this meeting).

V. Action items from previous meeting:

A. Feedback from Camilla Plata Corredor (Plinian Core) and David Fichtmueller (ABCD/EFG) following discussion of issues related to XML-based standards. (allow approximately 15 minutes).
Tim: I'm told a group will be submitting this for ratification in 2022. We will surely need a TAG response to this. I believe it will be a DwC-A extension for the most part
Camila: It will be several DwC-A extensions for the Taxon Core
Slides: https://docs.google.com/presentation/d/1Nnr2dQbBDlxDWEzIqg6YQLIka888sVwQIVpDOVR1tvU also preserved in the TAG repo (documents folder) at https://github.com/tdwg/tag/blob/master/documents/documenting_xml_standards_with_the_tdwg_sds_2022-06-27.pdf .
Conclusions:
- The SDS should apply to XML standards
- What is missing is structure and context. The structure can be defined by including an xsd-schema (or analogs RelaxNG or DTD) as a normative machine-readable document. For the human-readable, term-level documentation, the XPath can be used to relate conceptual terms to particular nodes in the schema.

- - To handle element reuse, there are two strategies: define a term and link it to multiple XPaths (preferable for Plinian Core) or define a term for each reuse and link each to a single XPath (preferable for ABCD).
    - There are some other complicating factors like mixed content (text+child nodes), loops (an element inside an instance of the same element), ordering, attributes, and cardinality. However, these are not required to accomplish the primary linking of defined terms to hierarchical position of corresponding node(s) in the schema.
  - Additional thoughts:
    - The proposed approach handles making an XML-based standard compliant with the SDS, but doesn't go the other direction (adding structure to an existing SDS-compliant vocabular).
    - It would be helpful to document the steps they took to move from the XML schema to "flat" documentation.
    - JSON-Path could be used as an analogous solution if there were a JSON Schema-based standard.
  - Camila noted that some terms in Plinian Core have many XPaths, so it might be desirable to make the Markdown/HTML have that field be collapsable so that viewers of the List of Terms don't have to see it unless they want to.
  B. Any followup from Raïssa Meyer and James Macklin regarding potential "champions" to advance GGBN through the TDWG process? Alternative course of action is to do nothing and let GGBN stand as it does now as a non-TDWG standard. Raïssa received a response on 2022-06-23 from Katharine Barker who is trying to identify an appropriate technical point of contact.
  James Macklin: Anton Guentsch replied that we should wait for Gabi's return, which is hoped to be later in summer.
  C. Rod Page unable to attend, but started a thread on the Slack channel and suggested reviewing the existing GUID and LSID Applicability Statements. Other comments or suggestions for moving forward? Potential participants in a Task Group were identified at last meeting as Rod Page, Jonathan Rees, Rob Sanderson, Ian Engelbrecht, possibly Larry Lannom from DiSSCO via James. See also https://github.com/tdwg/tag/issues/9 which has a 2017 list of people who were interested in the issue and action items. However, this issue never moved forward at that time.
VI. New item: *GBIF data model and its place in TDWG.* At a recent Executive Committee meeting, there was discussion about potentially having some kind of TDWG informational event about the developing GBIF data model (https://www.gbif.org/new-data-model). It was noted that this isn't actually a TDWG effort (although TDWG people are involved in developing it). The question came up about what role this model may play in the future of TDWG as sort of community domain model. Since that gets into the realm of the TAG, the Executive Committee asked that the TAG discuss and perhaps clarify what place the GBIF model may have with respect to ratified TDWG standards.
  A. John Wieczorek and Tim Robertson are the main developers of the model and are also TAG members, and therefore may be able to weigh in on this question.

- *Tim and John recommend TDWG take a position of observing, promoting engagement and giving feedback but that it is too soon to know where in TDWG this should land. GBIF anticipate this will result in the evolution of existing, or possibly a new standard(s) in TDWG at some point in the future.*

    B. If the model were added to an existing vocabulary standard (i.e. Darwin Core), its addition would be governed by the Vocabulary Maintenance Specification. It would be considered a "vocabulary enhancement" ([Section 1.4](#)), with the process being governed by [Section 4.1](#) . Note that a vocabulary enhancement can come from a group that "may arise spontaneously in response to community need"; it does not need to be a chartered Task Group. There would probably be a different pathway if it were considered to apply beyond Darwin Core – probably following the [TDWG Process](#) for new standards.

    C. Notes from comments during meeting: John W.:They are developing through field testing, but currently don't aspire to be creating a standard (yet). Working in an agile manner to develop solutions that work across a broad range of use cases. This gives them the freedom to develop something that works. After that, they can see if the community finds value in turning it into a standard. David S. question: are there terms that would strain the use of existing terms in DwC? Creating a more relational model is going to require minting more terms that link the various tables (used in the Frictionless package). At what point do we need to consider that those new terms become part of the standard? John: some terms have been used in two different ways that aren't compatible. That would require creating a new term that has the correct semantics. Steve: there is a precedent for this in the DwC RDF Guide, where linking terms were minted in cases where the relationships between classes were clear (see Table 3.6 in the RDF Guide for examples http://rs.tdwg.org/dwc/terms/guides/rdf/). David S.: what's the role of the TAG? Do we wait until the relationships gel to take action? Would the TAG review this? Steve: the answer depends partly on whether this is considered to be solely related to Darwin Core (in which case the DwC Maintenance Group would handle the term changes) or TDWG wide (including other standards such as ABCD, Latimer Core, etc.) in which case the TAG may need to play a role.

VII. New item: *OntoCommons/TDWG collaboration call*. The Executive Committee received an email from the OntoCommons group suggesting a collaboration between them and TDWG. On 2022-06-23 Steve and Theary Ung (TDWG Secretary) were on a call with them to learn more about what they had in mind. They provided a backgrounder using a slide set and the suggested areas of collaboration were summarized in another slide set. Part of what they were interested in was a better understanding of our standards development and management practices as represented in the Standards Documentation Specification (SDS) and Vocabulary Maintenance Specification (VMS) and Steve will follow up with them to answer questions about that. The other suggested points of collaboration involved more labor-intensive things, such as participation in a working group, speaking in a workshop, and serving as an expert in their "Standardisation Focus Area".

A. I stressed that TDWG's past experience with formal ontology development was not great (the TDWG Ontology over 10 years ago; see http://www.gbif.org/resource/80862 for historical background if you are unfamiliar with this), and that currently our vocabularies are very light-weight ontologies on the "bag of terms" level – very different from the approach they are taking.

B. I would like to gauge the level of interest in this and whether we can identify individuals who would have the bandwidth to get involved.

C. If there is sufficient interest, it might be beneficial to schedule another meeting with them that would include any interested parties from TDWG.
Steve has already indicated he's available to answer questions about the SDS, VMS, and TDWG maintenance processes. Jonathan Rees noted that this would be a lot of work, but offered to meet with them to have some initial conversations.

VIII. Any additional announcements or new business (include your name):

A. Has anyone else been approached to develop Biodiversity standards connected to ISO? I got asked to "apply" to be a member of a mirror committee of ISO via our Canadian Standards Organization.(James)
Tim Robertson: I forget the details but I recall a discussion around 10yrs ago in GBIF circles that ended with a "we need to commit to open discussion and open standards only" (my paraphrasing).
If ISO still has a pay-to-view approach to accessing standards I think TDWG should steer clear as it goes against the core mission and values.
Jonathan R. commented that he liked the idea of TDWG participating defensively, not necessarily being involved in the creation of ISO standards, but watching what they are doing to make sure that it's coordinated with TDWG standards. James will apply and let us know what they say. He will coordinate with Steve in making contact.

IX. Action items for next meeting (or before):

A. Steve will report to Executive that the TAG supports moving Latimer Core to the appointment of a review manager.

B. Ben Norton and Tim Robertson will make a recommendation for a TDWG-wide policy on the format of boolean values, with input from anyone who cares to give it. Tim mentioned the possibility of guidance on broader datatyping issues. That is optional, but we need a decision on booleans more immediately.

C. Steve will report John and Tim's statement about the GBIF model to the Executive and refer them to these notes to see the discussion that took place.

D. Steve will coordinate with Jonathan about making contact with the OntoCommons people and with Theary will report back to the Executive the details of our meeting with them.

E. Follow up on the status of creating a GUID Task Group (identify potential members and get a commitment from a convener or co-conveners).

F. Follow up on contacts with GGBN, in particular Gabi Dröge and whether she or someone else can step up to move the ratification forward or if we should just leave it as an external standard.