TDWG Technical Architecture Group (TAG)
Online meeting
2022-03-23 13:00 UTC

Meeting notes:

Participating:
- Steve Baskauf,
- Tim Robertson (first 30 mins only)
- David Fichtmueller
- Ian Engelbrecht
- James Macklin
- Ben Norton
- Jeff Gerbracht
- Roderic Page
- David Shorthouse
- Robert Sanderson (Yale University)
- Sarah Vincent
- Camila Plata

Regrets: John Wieczorek, Raïssa Meyer,
Debbie Paul (welcoming, lurking :D)

Notes taken during the meeting are in red.

I. Welcome
II. Set time and date for next meeting. Suggest: Wed May 11 at 13:00 UTC. Could be later. Note: John Wieczorek reports that this time would only work for him on Mondays, so 13:00 UTC on a Monday could be an alternative. Set next meeting time for 13:00 UTC, Monday, May 9.
III. Charter revisions.
    A. See 🗒 tag_charter_revisions for proposed changes.
    B. See https://github.com/tdwg/tag/commit/ac1740f9ee117bbd716fa3751f0b6e5f4e582e1d for diff and https://github.com/tdwg/tag/blob/master/documents/proposed_charter_chages.md for the revised draft.
    C. Decide whether to submit proposed changes to Executive. There was some discussion about the desirability of members receiving an agenda in a timely fashion so that they can prepare, but the consensus was that it was not necessary to specify a certain number of days. There was also some discussion about the quorum, where it was noted that this was helpful in some cases. But in the end, there was support for making the change based on the rationale given in the proposal.
IV. Overview of role of the TAG.
    A. Advise Executive when called upon
        1. review proposed task group charters
        2. technical review of proposed additions and changes

B. Proactive technical oversight
   1. completeness and interoperability of new and existing standards
      Comment from Tim Robertson: Does this cover aspects of a "common domain model" for TDWG? If not, should it?
   2. interaction with external standards activity
C. Commons standards framework
   1. coordination of technical activity across Interest Groups
   2. maintain "road map" and best practices documentation
D. Historical lessons
E. Some comments made by Rob Sanderson about when a TAG can play an important role: 1 - Before approving a new working group / standard for development. e.g. Is this technically valuable / feasible. 2 - For consultation during development, on demand. This is super useful in the W3C, where folks who have the best overview of everything going on can answer tricky cross-spec questions. And 3 - Before public review. Is it technically good enough to ask non member / non experts to review?
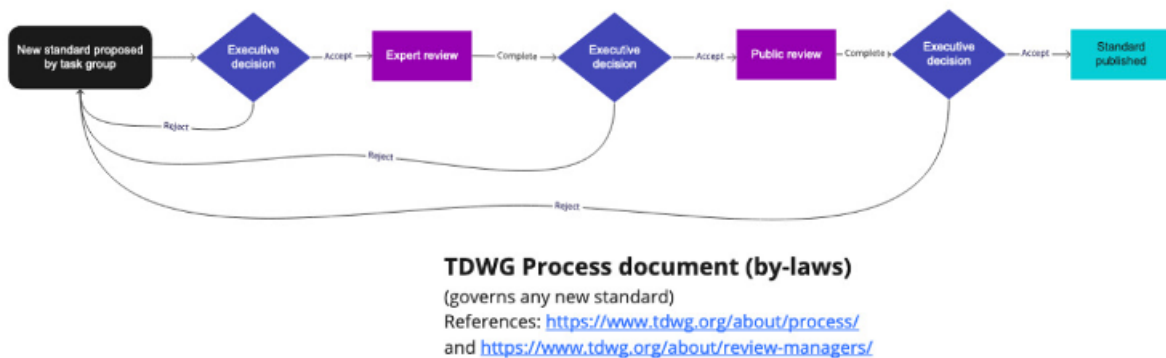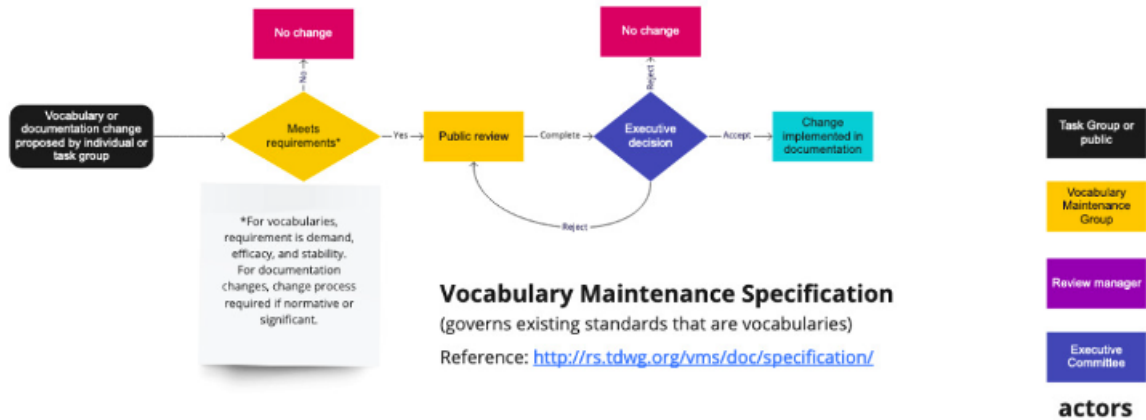
V. Review of TDWG administrative practices.
   A. Governance of standards process

	vocabulary	non-vocabulary
new standard	TDWG Process (by-laws), guided by a Review Manager	
change to a standard	Vocabulary Maintenance Specification (VMS), guided by a Maintenance Group	no established process

All TDWG standards are expected to conform to the Standards Documentation Specification (SDS). However, there are some specific requirements for vocabularies that do not apply to other standards.

   1. Refer to http://rs.tdwg.org/vms/doc/specification/ https://www.tdwg.org/about/process/ and https://www.tdwg.org/about/review-managers/ Comment from Tim Robertson in doc: Time for HTTPS universally? Later response from Steve B.: one big issue to be considered is that those IRIs are also being used as permanent, stable identifiers. So despite the desirability of using HTTPS for URL resolution, adding the "s" would effectively change all of the identifiers. Other organizations have the same problem (Dublin Core, schema.org, FOAF) and haven't changed the identifiers, presumably for this reason.
   B. Workflows

Vocabulary Maintenance Specification
(governs existing standards that are vocabularies)
Reference: http://rs.tdwg.org/vms/doc/specification/

*For vocabularies, requirement is demand, efficacy, and stability. For documentation changes, change process required if normative or significant.

actors

TDWG Process document (by-laws)
(governs any new standard)
References: https://www.tdwg.org/about/process/
and https://www.tdwg.org/about/review-managers/

1. See https://github.com/tdwg/tag/blob/master/documents/tdwg_change_processes.pdf for higher quality image.
2. Note: TDWG Process directed by a review manager and VMS process directed by a Vocabulary Maintenance Group

VI. Status of current standards, draft standards, and coordinated additions to standards.
    A. See https://docs.google.com/spreadsheets/d/1qYWbMPaQMehfYT2txUehCFe77l-NMczyDQkqqc0eVv0/edit?usp=sharing compiled from 2021 annual reports to the Executive.
VII. Key issues identified
    A. Issues with XML-based standards
        1. Involves ABCD, EFG, and Plinian Core
        2. Is an XML schema-based standard a vocabulary?
            a) If so, subject to vocabulary prescriptions in SDS (bag of terms, basic metadata)
            b) If not, what are the guidelines?
        3. What are the documentation requirements?
            a) Cover page and schema doc satisfactory?
            b) Is a "list of terms" document required?
                (1) If so, what is a "term"?
                (2) If not, how are changes managed, IRIs dereferenced, etc. (rs.tdwg.org repo vs. BioWikiFarm)?
    B. Identifier-based issues

1. How to handle the ggbn: namespace terms in the GGBN proposal (https://github.com/tdwg/ggbn/issues/3)
2. Should the LSID applicability statement be retired? What would be the implications?
3. Should the GUID applicability statement be revised?
    a) Who are the stakeholders? Current practices of CETAF (https://cetaf.org/)
    b) Potential fait accompli if GBIF or BiCIKL start issuing identifiers. Note from Tim Robertson: BiCIKL is a project involving several infrastructures, not an infrastructure itself.
4. Role of Wikidata Q IDs and Wikidata as "identifier central"
C. Modelling and serialization issues
    1. DwC/ABCD consensus model development stalled.
    2. Star schema replacement effort at GBIF under development and will effectively result in a DwC data model. Note from Tim R.: It has the potential to be a model to encompass things that draw from more than DwC (Audubon Core, Humboldt Core, ABCD etc)
    3. Latimer Core (Collections Description) has a complex model – how will it be serialized?
    4. The Humboldt Extension to DwC has a somewhat complex nested event model, compatible with the IPT. Note from John Wieczorek: The extension (see http://rs.gbif.org/sandbox/extension/Humboldt_2020-12-11.xml) does not have any nesting, but is meant to extend the Event Core, which does enable nesting via the parentEventID. Note from Tim R.: GBIF and ALA now working on SamplingEvent pipelines that will index Event records, in addition to occurrences (early days still, but see https://docs.google.com/document/d/19ZXWKgyFBexNXI8sN-sjizEB1eHjpHRTxeC05zf2MbU/edit)
    5. Do we actually aspire to:
        a) have a single serialization paradigm that can operate across standards (JSON-LD?) Note from Tim R.: We should be cautious here. Guidelines for serialization may be in scope but forcing one should not be. Look at Avro, Parquet, Protobuf, Apache Arrow etc and see the state of modern data platforms that use far more than JSON. See e.g. https://data-blog.gbif.org/post/apache-arrow-and-parquet/ (Arrow/Parquet) and https://registry.opendata.aws/gbif/ (Parquet). The important aspect should be schemas, the concepts they refer to and schema evolution, not the serialization format in my opinion.
        b) break down silos by supporting Linked Data (RDF)?
        c) coordinate with external standards (W3C Annotations? Frictionless Data? Non-TDWG data standards organizations) Tim R comment: I believe, yes to some extent. E.g. IPT v3 of GBIF exploring FD. https://github.com/gbif/ipt/issues/1736 Keeping an eye on this stuff will help ensure our

interoperability and ensure we are current and relevant. TAG has a role to play in helping advise TDWG on these aspects.

VIII. Action items. Possible actions by: individual TAG members, groups led by TAG members, external groups with TAG member participation, external consultants

    A. XML issues

        1. The EFG publication issue MUST be solved. There was discussion in the documents comments asking about this. The basic problem was that EFG went through the review process with 3 expert reviewers and was submitted for ratification before the SDS came into effect. However, the public review was not done until 2020, after the SDS. In fairness to the authors and because EFG is already in use, it was ratified by the Executive Committee without it being in conformance to the SDS. So it now needs to be "fixed" retroactively.

    B. Identifier issues

        1. GGBN Task Group doesn't have specific representation in the TAG, needs a champion.

    C. Modelling/serialization issues