

Agenda/Meeting notes:

Participating:

- Steve Baskauf
- David Fichtmueller
- Ben Norton
- Sarah Vincent
- Kit Lewers
- James Macklin
- Ian Engelbrecht
- David Shorthouse
- Roderic Page
- Greg Whitbread
- Holly Little
- John Wieczorek

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.

As with previous meetings, Steve will record the meeting for note taking and later viewing.

- I. Set time and date for next meeting. 2023-03-13 13:00 UTC.
- II. Change in group composition for 2023:
 - A. Jeff Gerbracht leaving as an official member.
 - B. Kit Lewers joining. Kit's interest is information science and is at U. Colorado.
She has agreed to take an organizational role in reviewing prior standards.
- III. Follow-up on items from the previous meeting
 - A. GUID organization meeting to be scheduled for early 2023 (is there a date set for this, Ian?)
 1. Date not set yet but will do so in the next week or so.
 - B. Policy on boolean values.
 1. Draft controlled vocabulary still at:
<https://github.com/tdwg/tag/blob/master/build/boolean/index.md>
 2. Consensus at previous meeting was that such recommendations don't need to go through the formal ratification process. Since the TAG should be monitoring standards development, if people don't follow the TAG recommendations, they should be asked to provide a reason why. If there isn't a good reason, they should follow the recommendations. There has been discussion among the team

revising the website to create a place for recommendations and best practices. So we can put policies there when it gets fleshed out.

3. Best practices document at <https://github.com/tdwg/tag/blob/master/boolean/boolean-best-practice.s.md> Designed to complement Ben's reference document at <https://docs.google.com/document/d/14f7H7hSTyiu0LXCJc9nG-uvy3-ElAkUd/edit?usp=sharing&oid=116355305033345986500&rtpof=true&sd=true>
4. Critical question: should data transfer formats that allow datatyping (XML, JSON) serialize booleans as datatyped values rather than the plain strings "true" and "false"? Do we do that for other things that can be datatyped like integers when output from an API in JSON, e.g. {"individualCount": 3} and not {"individualCount": "3"} ? This is somewhat of a deviation from the spirit of Section 1.4 of the VMS: <https://github.com/tdwg/vocab/blob/master/vms/maintenance-specification.md#14-the-general-structure-of-tdwg-vocabularies>, which says that terms should have few constraints with regards to datatyping. John Wieczorek noted that he sees these recommendations as being a layer over the "bag of terms" so he doesn't see a deviation in that sense.
5. Discussion: The GBIF API responds in various ways. dwc:decimalLatitude and dwc:decimalLongitude are expressed as floats. dwc:individualCount returns strings, but that may be to accommodate ranges like "1-2". GBIF expresses their own term "isInCluster" as boolean. Rod Page says that he favors JSON typing for boolean, integers, and floats. iNaturalist follows JSON typing for integers, floats, and booleans. John Wieczorek notes that GBIF follows JSON typing when they present interpreted data over which they control the content and therefore the datatype. This is an argument for the verbatim terms where the original data may be presented as strings. Ben Norton notes that applications that output data have to cast them as a particular type (e.g. integer). Since people are sloppy you find output both ways (typed or as strings). If we say that you should cast booleans without quotations, we will still get both. Steve notes that although we can't enforce RFC 2119 MUST statements, it's a first step towards encouraging people to be more consistent. Ben: Awareness will help a lot, people would be less likely to overlook. The consensus seems to be that if it's possible to indicate that a value is a boolean, it should be, by whatever means are available in a particular serialization system.

C. Recommendations for expressing complex values

1. See discussion in red under item VI. in last meeting notes for background. https://docs.google.com/document/d/1RPwmyXHpsD5xPNrKWLO_qqCWBb_8npEK660PvwyVII8/edit?usp=sharing

2. How to move this forward? Chartering a task group is probably overkill, but we may need a subcommittee to write proposed recommendations. In February, Ben Norton will take the lead on writing this up. John Wieczorek offered to support Ben in it as necessary.

IV. New items

- A. Review standards status and recommend standards to be retired. Kit Lewers has volunteered to take administrative responsibility for this.

1. There is a new, single issue for this:

<https://github.com/tdwg/tag/issues/37>

Kit has started reaching out to authors of the old standards to inquire about how much they are being used.

- B. Best practices for borrowing terms from non-TDWG vocabularies.

1. See <https://github.com/tdwg/tag/issues/39> for details.

2. If we have any feedback about this, we should give it prior to the 2022-01-18 Material Sample Task Group meeting.

Discussion: Daved Shorthouse notes that in his opinion, BFO has terrible labels (ed: local names) that are counter-productive for humans that use them (e.g. for mapping in IPT). He tried using BFO for agent roles in an AgentActions extension to DwC-A and it did not work well.

John W. noted that on the IPT, the problem could be overcome by using labels rather than the term's short name (i.e. local name). Thus issues with the IPT shouldn't necessarily form the basis of a technical decision.

David S response: Fair enough, but are there examples in our community where mapping is done prior to publishing data where labels are used as opposed to a short name? I'm not aware of one. In any case, the technical recommendation here will incur a significant development hit across the board.

Steve B. noted that in Audubon Core terms were not minted when there was a possibility of borrowing existing external terms. A "usage" metadata property (skos:scopeNote) was used to explain the specifics of how AC intended for the term to be used in that particular context.

Sara Vincent: Another approach (used in Latimer Core for example) is to mint new terms and use SKOS mappings (broader, closeMatch, etc.) to indicate how the minted terms relate to external terms.

https://github.com/tdwg/cd/blob/review/standard/terms/ltc_skos_mapping.csv

John W. such mappings have been made formally between MlxS and DwC terms:

<https://github.com/tdwg/gbwg/tree/main/dwc-mixs/mapping> .

Ben Norton: In Latimer Core, their criterion for borrowing a term rather than minting is that the borrowed term must be a skos:exactMatch to be eligible to be borrowed. That is a high bar. The mapping table approach works well when the desired use for the term is narrower than the external term.

John W. mentions an instance where a DwC term was deprecated in favor of a more well-known external term: `dwc:source` deprecated in favor of `dcterms:references` .

Ian Engelbrecht: It does seem like this is quite an important issue – it's an example that might set a precedent. As the TAG, we probably should be providing some input on that rather than letting it fall back on a task group. He wonders if we have enough time here to give this the attention it needs or if it needs more thought and unpacking before we actually make a decision and recommendation.

Steve: If we feel we need to weigh in on this, then we need to communicate with the MaterialSample task group and tell them. They can come to our meetings and weigh in. This is closely related to the whole mapping issue that we are discussing.

Sarah V.: Would it be a feasible recommendation to do a SKOS mapping and then call the term whatever they want? Because that seems to be what some other standards are doing. If in the future the TAG recommends that the SKOS mapping documentation becomes a part of the official normative documentation, then you can call your new term whatever makes sense. And everything that has a SKOS mapping to it has its relationship expressed in a standardized way. So anyone who wants to look at those linkages can do so.

Steve B. mentions how Version 2 of the W3C DCAT recommendation changes some of its terms from Version 1 to remove constraints. Note added later: see <https://www.w3.org/TR/vocab-dcat-2/> "Status of This Document". The particular case was superclassing of `dc:Dataset`. In V1, `dc:Dataset` was a subclass of `dctype:Dataset` . In V2, the superclass is `dc:Resource`, which is much broader than `dctype:Dataset`.

Steve notes that perhaps the solution is to just mint `dwc:MaterialEntity` and declare it to be a subclass of `dcterms:PhysicalResource` .

John W. Can we come up with a short set of conditions under which borrowing is recommended over minting? Can we ask MaterialSample to look at the problem this way?

Recommended course of action:

1. Seed the TAG Slack with the Latimer Core criterion for deciding whether to borrow or mint and continue the discussion offline. If something concrete happens there between now and the MaterialSample TG meeting next week, we can take that to them as a recommendation.
2. TAG members who have a long view of how this issue has played out in the past should attend the MaterialSample meeting and contribute to the discussion.
3. As the MaterialSample group decides on this particular proposal, they should also articulate the "short set of conditions" they used to decide their recommendations. That can be developed into a precedent/policy.

C. Path forward for Standards Mappings. Identifying potential core members and convener. Points from previous meeting:

1. Mapping standard SSOM was mentioned.
2. There is now a channel in the regular TDWG Slack: #mappings-between-standards for discussion.
3. Holly Little suggested using the paleo group as a use case because there are a lot of different standards in that field.

David Fichtmueller suggested that he and Holly might organize an initial meeting of people interested in standards mapping and find someone in that group to serve as a convener. It would probably be no sooner than the second half of February.

V. Any additional announcements or new business.

None

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

- A. Members should take a careful look at the boolean list of terms and best practices documents, and suggest changes via pull request. Ben will review his boolean reference document to potentially make it fit together better with the other two new documents. Our goal by the March meeting would be to have a draft that we could recommend for posting on the TDWG website.
- B. In February, Ben Norton will work on a draft of recommendations and examples for providing complex values (e.g. lists). John Wieczorek is available for support in this task and any others who want to contribute should contact Ben.
- C. Continue discussion about criteria for borrowing vs. minting and mapping on Slack over the next week. Members attend MaterialSample meeting and provide input to that task group.
- D. David Fichtmueller (and possibly Holly Little) will organize a meeting of people interested in standards mapping to find a convener and core members for a task group (no earlier than 3rd week of February).