Darwin Core Hour Notes - 2017
with link to 2018 Notes

Table of Contents

(each one links to the actual spot on this doc, click on it and follow the link)

Darwin Core Hour 7 Feb 2017 (WEBINAR)- Notes

Calendar Announcement https://www.idigbio.org/content/darwin-core-hour-webinar-series
Darwin Core Input Form: https://tinyurl.com/zja2muz
Post Webinar Survey: https://tinyurl.com/jq8qzsq/
Recording: http://idigbio.adobeconnect.com/p200obyl8yp/
Vimeo:

Please capture notes relevant to next steps / to-do's / ideas that come up. Thanks!

Darwin Core T-shirt ($19.99 - see D.Bloom) *would need a DwC logo, no…? ;)
Audubon Core on the back!

People could not join because attendees limit was reached! Will need some solution for next webinars. Some forms submitted about this.
One idea: People can room-share with colleagues to save log-ins for others. (PZ:)Like this idea, but should maybe be told in advance somewhere, maybe on the remainder of the webinar?

DarwinCore Documentation (inspired by the 2016 TDWG meeting presentation by Paula Zermoglio)
https://tinyurl.com/jcozm3l Paula's recorded presentation - go to time offset 0:08:25
Additional suggestions for discussion topics should be submitted to the DarwinCore Q&A GitHub site
http://github.com/tdwg/dwc-qa

Brian: How would I get on that list John just mentioned? [//from Deb - I don't know what Brian's referring to here - I was busy navigating emails of "help I can't get in"]
[Deb, Brian wanted to join the Biodiversity Informatics Training Curriculum group
Thanks Gary - so we need to (or did someone) get Brian the BITC link?
Yes, Erica did that.
Thanks Erica and Gary!
https://plus.google.com/communities/111802729072058850441
http://biodiversity-informatics-training.org/ (this is the link that Erica shared) ← EK: we should share the direct link to the google group above

HeatherC @AAFC 2: would be great if the Darwin Core term documentation linked to the relevant GitHub "issues" to further inform and guide users
RE: that is the idea. Improve documentation (right now not available, not complete) and link it to the questions/issues.

Darwin Core Terms - http://rs.tdwg.org/dwc/terms/

SophieP: Hello! Is versioning of the terms planned on the GitHub page and/or DwC wiki page? (question asked a lot of time during trainings ;))

RE: discuss this? Versioning is on tdwg pages -history of terms. Last versioning is always the one on the terms page. This is shown in presentation as well. Also discussed by JW in questions time.

Feedback from Town Peterson //inserted here by Deb:

> Hi all,
> OK, so I have managed to sneak into the DwC hour. Good that you guys are getting this going. Anything and everything that I/we can do to help ... here are some comments...
>
> 1. Doing it at 1 pm Eastern will essentially cut out anyone in Europe, Africa, or Asia. We found that midmorning eastern time works better, given that the Western Pacific is mostly empty.
>
> 2. Adobe Connect has two limitations that are highly relevant ... (1) it has a connect limit, as you found on this first broadcast, which is presently excluding my lab group, and (2) it requires a software install, which means that anyone who must watch in a public facility or a facility where she/he does not have install rights. THere was also some breakup of the sound, around the "Where did DwC come from" slide that John put up.
>
> 3. A content suggestion ... and one in which I would be willing to help. A presentation on the overwhelming / critical nature of CoordinatePrecisionInMeters as an enabling field. This is little understood by the user community, and indeed that field was not even SERVED by GBIF until last year! Bridging between the informatics side and the user side in a single presentation could be very informative. Just a thought.
>
> Anyhow, those are some first thoughts ... All the best, ATP

Normative Darwin Core

- Darwin Core Quick Reference Guide - http://rs.tdwg.org/dwc/terms/
  - TDWG Semantic MediaWiki - http://terms.tdwg.org/wiki/Darwin_Core (John not certain about being able to use this for new translations/info going forward, likely but not 100%)

Darwin Core (DwC) + Dublin Core (DC) + Audubon Core (AC)
Biodiversity resource terms + resource terms + biodiversity media resource terms

Darwin Day Feb 12!

DwC Terms: Vocabularies. Classes, Properties, -Values-.

- Metadata

Datasets--> Metadata--> Records--> Fields--> Values.

s: Do you declare your vocabularies on the same darwin core field?
RE: one populates the fields only with values, one does not explicitly declare the vocabs in the same fields where the actual data is being captured.
S: do you declare the vocabularies at the head when you declare your schema? or do you wrap each data element with the schema dwc and the property controlled vocab used?

(EK): is there a way to declare this?? In the EML metadata file on a DWC archive? I also share this question.

Simple DwC and/vs. **DwC extensions** (for non 1:1 relationships to simple DwC)

- GBIF extensions site *http://tools.gbif.org/dwca-validator/extensions.do) with caution, to be changed soon.
- Christian Gendreau (GBIF): The information on the Extensions page will be preserved, it will simply moved to a new page.
- Sandra: Are there extensions related to trophic interactions or species predation habits or food?
  RE: Deb Paul: Sandra, folks are indeed working on standards for sharing species interactions, habits, traits data
- AlexH: Where is Event Core extension to be found?
  RE: David Bloom: AlexH - there is a link via GBIF, http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Event

TDWG Biodiversity Information Standards Group

**Need to send link of the recording to BID community and others - on vimeo for download; ppt requested also http://idigbio.adobeconnect.com/p200obyl8yp/

Jonathan Amith: I have a question related to ethnobiology. The questions fall into two categories. The first concerns limitations of DwC for ingesting and discovering datasets that are of interest to ethnobiologists. The second concerns possible extensions to DwC for linguistic/cultural information. (On similar note: Zooarchaeology Database being developed at FLMNH)

Xiaoli Ma: What are the tools in the end the community will produce after improving Darwin Core?

Brian: Pretty basic, but could you differentiate DwC from DwC Archive?
RE: Erica Krimmel: @Brian, DwC are the rules and a DwC archive is data/metadata generated according to those rules
RE: Laura Russell: Brian, DwC is the Standard, DwCA is the text output mapping the standard with the actual data. http://rs.tdwg.org/dwc/terms/guides/text/index.htm

Data Mgmt Interest Group - Meeting 14 Feb 2017

Present: Joanna McCaffrey, Shelley James, Paula Zermoglio, John Wieczorek, Erica Krimmel, Rich Rabeler, Deborah Paul, Mare Nazaire

- DwC Hour Workflow - mgmt of our process - Paula's vision
  https://docs.google.com/drawings/d/1vnDyJB5zl7HS1LZd7OFHq6ZGl5iZdhNEK3ytwbFVtlA/edit

- DwC Hour itself: presentation > discussion > 10 minute section at the end to review current tickets / requests (as a way to answer some questions in a more timely manner)
- Each DwC Hour, each DwC Announcement, needs:
    - info about upcoming new DwC Hours
    - link to old DwC hour
    - link to FAQ *in development*
    - link to where community can review status of pending issues from prior meetings (GitHub)
    - ?show the workflow https://docs.google.com/drawings/d/1vnDyJB5zl7HS1LZd7OFHq6ZGl5iZdhNEK3ytwbFVtlA/edit
- John Wieczorek to ping Kevin to ask
    - please show how connection from google sheet to github repo is done
    - want to make a label “form submission” for each github ticket that comes in via the form
        - is this done literally in the script?
        - would it help for Deb to add a field to the Google Form that declares this github label literally?
    - Deb is going to add a Please share your github username if you have one (here’s where to get one if you don’t have one). We’ll use this to make sure we can contact you to answer your question.
- Deb to send email to those offering to help cover topics - when can they be ready to present something?

Next DwC Hour Webinars:

A. **Darwin Core Hour #2** - March 7th, 11 AM EST, 4 PM UTC, 1 PM Buenos Aires - John Wieczorek

a. Controlled Vocabularies and where they come from

i. key examples: dcterms:type, dwc:basisOfRecord
    1. ?exicatti

ii. issues

iii. what fields have or are suggested to have controlled vocabs (that may / can / will vary across communities (See list: https://docs.google.com/document/d/1IlEbOypxX6pnoAZY0pjWcl4MhsIN5K85KDNqKxczTCg/edit)

iv. what data can we see inside these fields for this from
    1. GBIF
    2. iDigBio
    3. VertNet
    4. ALA maybe

v. use a repeatable query where we can use this as a community metric. Is the data (1 year later) becoming more standard, for example.

Title: Darwin Core Hour: Even Simple is Hard

Abstract:

In this chapter of the Darwin Core Hour series, we will glance at the list of Darwin Core terms that are recommended to be populated with values from controlled vocabularies. From that list we’ll look at some basic terms (e.g., dcterms:type, basisOfRecord, etc.) for which there are clearly recommended controlled vocabularies. We will explore how they are used in practice and how usage differs from the recommendations. We’ll discuss what the consequences might be of not following controlled vocabularies, and the converse - what the consequences might be of doing so. We’ll look at some of the secrets of how aggregators deal with the problem and the

lessons we can learn from them. Finally, we will review questions that have been submitted related to this particular subject and open the webinar for further discussion.

B. **Darwin Core Hour #3** - April 4th, 11 AM EDT, 4 PM UTC, 12 Buenos Aires - Paula Zermoglio
   Title: Darwin Core Hour: Thousands of shades for "Controlled" Vocabularies
   Abstract:
   In this third chapter in the Darwin Core Hour series we will visit some of the most colorful Darwin Core terms for which the use of controlled vocabularies is recommended. First, as a follow up to Chapter 2, we will expose the current content of particular terms as they are published right now via different aggregators. Second, we will try to disentangle the reasons, purposes and chances that lead us to observe this diversity of values in such fields. Then, we will explore the current availability of controlled vocabularies in different disciplines within our community and the initiatives that are addressing the problem from different perspectives. Finally, we will try to understand if there is actually a pot at the end of the rainbow: can we come up with solid, community-built controlled vocabularies?

C. Darwin Core Hour #4
D. Darwin Core Hour #5
E. Darwin Core Hour #6 - July 11? Georeferencing Terms in Darwin Core
   a. dwc georeferencing fields -- locality specific
   b. Not! how to georeference, rather, what does / doesn't go in particular fields
   c.
F.
G. Darwin Core Hour #7
H. Darwin Core Hour #8

Future DwC Hour Webinars

- dwc controlled vocabularies: taxonomy
- dwc controlled vocabularies: geography
- dwc controlled vocabularies: Examples and Community Solutions (continuation of DwC hours' 2 and 3)
- dwc archives -- dwc extensions, GBIF, David Shorthouse, Canadensys
- dwc publishing -- dwc archives via IPT -- Joanna McCaffrey, Laura Russell, Kyle Braak
- dwc aggregator specific presentations (VertNet, iDigBio, GBIF, ...)
- dwc data quality strategies -- share / discover / develop - ask TDWG-GBIF DQ group to take part
- dwc kurator workflows
- dwc - audubon core -- image / media data / 3-D -- Gary Motz (said yes - is preparing)

Use "Special Guest" DwC Hours as needed when someone cannot make the repeated time.
Currently we seem to be the **First Tuesday of Each Month.**

Darwin Core Hour 7 Mar 2017 (WEBINAR)- Notes

Calendar Announcement https://www.idigbio.org/content/darwin-core-hour-even-simple-hard
Darwin Core Input Form: https://tinyurl.com/zja2muz
Post Webinar Survey: https://tinyurl.com/jxgk3uv/

Adobe Connect Recording: http://idigbio.adobeconnect.com/p2ewna3h59c/
Vimeo (mp4):
GitHub /tdwg/dwc-qa website: https://github.com/tdwg/dwc-qa
How to Navigate GitHub Darwin Core Hour Repository: https://youtu.be/saNUtQijRYM

Intro: dwc hour #2

Abstract:
In this 2nd chapter of the Darwin Core Hour series, we will glance **at the list of Darwin Core terms that are recommended to be populated with values from *controlled vocabularies.*** From that list we'll look at some basic terms (e.g., **dcterms:**type, **dwc**:basisOfRecord, etc.) for which there are clearly recommended controlled vocabularies. We will explore how they are used in practice and how usage differs from the recommendations. We'll discuss what the consequences might be of not following controlled vocabularies, and the converse - what the consequences might be of doing so. We'll look at some of the secrets of how aggregators deal with the problem and the lessons we can learn from them. Finally, we will review questions that have been submitted related to this particular subject and open the webinar for further discussion.

Presentation
John gave a pretty detailed overview of the DwC webinar series, the GitHub repository for tracking the Q&A submitted to the iDigBio Data Management Interest Group. Major kudos to Paula for documentation, labeling, and an especially excellent video tutorial on how to navigate the repository and GitHub, in general (https://youtu.be/saNUtQijRYM).

DwC "Terms". Class-properties-values
Potential confusion between names of classes also being names of values (e.g. PreservedSpecimen). Why, some "historical" reason, relationship between DwC - DublinC - AudubonCore (more specific vocabulary for media in biodiversity data)
Terms with recommended controlled vocabs (as in DwC -TDWG)

Type. Recommended vocab for it (DublinC vocab), which suitable for DwC.
Distinct values for Type found in GBIF.
basisOfRecord. Recommended vocab for it (DwC classes).
Distinct values for basisOfRecord found in GBIF. How many "interpretable"

By lack of reconciliation of standardized terms, we are precluding "understanding" by broad audiences (e.g. those who use DC but not necessarily DwC) and data aggregators help to maintain broad understanding of non-standard terms.

How can these "standard" terms be cleaned? The community MUST provide feedback to elaborate from initial simple and stable term usage. If terms like "basisOfRecord" are to allow for more specific standard values (e.g. herbariumSheet, tissueSample, fieldNotes, publication), the community must propose these values and advocate for broad adoption of the specific values for a DwC term
HerbariumSheet is not part of the standard and is useful for human readers, but creates lots of work for data aggregators and 'confuses' machines that are data mining. Your records can be rendered invisible if you don't conform to standard vocabularies and controlled values for DwC terms.

How could semantics help? Definition of terms AND relationships.

What can be done to improve data? Required fields, value compliance, stronger recommendations, controlled vocabs, better examples. GBIF Data Quality Requirements as an example.
http://www.gbif.org/publishing-data/quality

QUESTIONS: (see in orange the actual questions, see in black some answers and input)

1. Erica Clites (UCMP): As a data publisher, though, there are so many fields we could fill in. Can you give a clearer explanation for why we should send this field [preservedSpecimen]? How does not having this field inhibit research?

David S: May not inhibit research, but may inhibit reuse outside our field.

Talia & Virginia: Erica... having FossilSpecimen in basis of record makes it really easy to pull only fossil records when searching.

1.2. Erica Clites (UCMP): Exactly, so I'm just wondering if it is necessary to fill in dcterms:type if we are sending basisOfRecord.

David S: dcterms:type *should* be a window to another community of users, but maybe we can get John to provide some examples of reuse of DwC data outside of known biodiversity aggregators (eg librarians)

John (speaking, paraphrased): Did the presentation answer this already, Erica?

1.3. Erica Clites (UCMP): Not really, it seems like from what you said that it is interpretable by GBIF and they could fill it in.

John (speaking, paraphrased): Yes they could if we as a community did some pre-work.

FOLLOW UP

Elizabeth Martin: I think we need the Dublin Core type if we are hoping to share with those outside of the biological community since they may not know what a specimen is.

Brad Millen: I would expect the *Specimen is the field preparations

Erica Clites (UCMP): Maybe, I guess the answer is that it is helpful to people outside our field.

Brian: In "librarian-speak" it's a matter of "Interoperability"

Brad Millen: PInned Specimen, Herbarium Sheet are all preparations

GailK: Makes sense but it probably also represents a PreservedSpecimen

1.4. Deb Paul: dwc:basisOfRecord is then a subset of what's declared as dcterms:type

1.5. Dean Pentcheff (LACM): So... the determination is based on the person reading the text, not the observation type that's recorded in the text (?)

2. GailK: What if we are recording something from the literature (usually types), which we have not seen (for BasisOfRecord)?

GailK: I guess we are presuming these are PreservedSpecimen??

David S: GailK - in case of type info lifted from a publication, I see here that we'd consider this a HumanObservation

David S: GailK and the dcterms:type would be "Text"

3. William Ulate: So, does GBIF then pre-process datasets to conform to the used vocabulary? Is that the recommended procedure?

Sophie Pamerlon: William : Using the GBIF IPT, you cannot publish a dataset for which the basisOfRecord field is not filled in with a class from the controlled vocabulary

Sophie Pamerlon: and it is possible to assign a basisOfRecord to a whole dataset (even if the field does not exist in the source file) when doing the mapping with the IPT

John Wieczorek: GBIF requirements: http://www.gbif.org/publishing-data/quality

Sophie Pamerlon: (for required and recommended DwC terms on GBIF)

Joanna - iDigBio: iDigBio screens for basisOfRecord compliance

Matthew Collins: You can see the distribution in iDigBio: http://search.idigbio.org/v2/summary/top/records?top_fields=basisofrecord

Christian Gendreau (GBIF): We do try to interpret basisOfRecord in order to index them properly but the verbatim value is preserved

Matthew Collins: Oh, yes, we do a bit of interpretation, here you can look at the raw: http://search.idigbio.org/v2/summary/top/records?top_fields=data.dwc:basisOfRecord

Matthew Collins: But Joanna has done a great job of guiding providers to send us stuff that fits the vocabulary.

4. nestorbeltran: We have jaguar (P. onca) tracks as evidence of a biological record, there's no specimen and there was no direct observation. Any advice on which basisOfRecord and type should we use?

Christian Gendreau (GBIF): I would say probably MachineObservation

Deb Paul: Nestor - was is a human who reported seeing the tracks?

Joanna - iDigBio: Nestor: is the thing you are cataloging the photo of the track - MachineObservation, or a HumanObservation of a sighting?

nestorbeltran: A reliable forest ranger wrote down the coordinates of the places where he saw the tracks, but there is no picture of them.

David S: dcterms:type is likely Event

Deb Paul: and this would be HumanObservation

Joanna - iDigBio: sounds like HumanObservation then, like the sighting of a bird.

Mary Beth: How about "anecdotal"?

Christian Gendreau (GBIF): Example from Bird tracking - GPS tracking of Lesser Black-backed Gulls and Herring Gulls breeding at the southern North Sea coast: http://www.gbif.org/occurrence/1209707560/verbatim

Brad Millen: anecdotal implies no evidence.

Christian Gendreau (GBIF): (sorry my example is about GPS tracking, not tracks in the sense of footprint)

David S: footprint, scat, loose feathers, etc. are interesting

nestorbeltran: HumanObservation and Event makes sense, thanks a lot

4.1. David S: re: nestor's q - where does fact that HumanObservation is a footprint go?

Elizabeth Martin: re: David S and nestor's q. Maybe under occurrenceRemarks? Also some of the information could be included with the descriptive EML metadata record.

nestorbeltran: I think that the fact that HumanObservation is a footprint may go to OccurrenceRemarks?

nestorbeltran: Elizabeth +1

5. GailK: Enforcing Value Compliance will be problematic for seeing where changes/growth need to come

John (speaking, paraphrased): if we have DwC requirements in place we may be limiting data publishing and/or accidentally losing information by “cleaning” data to comply.

GailK: exactly (see Enforcing Value Compliance)

Ben Anhalt: Does enforcing value compliance on required fields create the danger of "alternative facts" where users are incentivized to randomly make a choice just to make the system happy?

6. Dean Pentcheff (LACM): Examples are critical. Those of you (like John) who live/breathe/eat this have the mental context of many interpretations to guide usage. Those of us jumping in occasionally have such a constrained perspective that we can easily make decisions that don't make sense in the larger context. [from Deb - Exactly - and we need to get Dean's permission to quote him!].

Jim: I agree with Dean. A broad array of examples, especially for different collection types (e.g., herbaria, insects, etc.), are critical.

Dean Pentcheff (LACM): Thanks (on examples)! The "right" answers are often immediately obvious to the Darwin Core pros, whereas those of us with a single perspective are prone to misunderstandings.

7. Sandra Brantley: Perhaps we should also take into account the level of training of the folks doing data entry and/or record checking. In small museums, overseeing all this is a trick.

8. Jonathan Amith: I have a question and will write it but maybe better in voice. What happens when a controlled vocabulary, let us say "Music" is valid for 10 years, let's say, but then the working group or its equivalent decides to split into two, e.g., "Instrumental music" and "Voice music". Thus a content before a certain date is different than after., So, let us say DwC adds to PreservedSpecimen "Hebariumsheet" and PinnedSpecimen" Are such types of splits acceptable

Jonathan Amith: OK. I gave a bad example, sorry, but the general question is still there, what happens when distinctions are introduced in a given field after a certain date.

Jonathan Amith: It seems my mic is not working, apologies. But would it be possible to answer the general question. Does it ever occur that a new controlled vocab term is introduced that splits or further defines already existing terms and thus terms in the controlled vocab are "time sensitive". Apologies for lack of sound

David S: Jonathan Amith's q is about the impact of versioning in controlled vocabularies

9. Deb Paul: Do we need a subcategory for basisOfRecord? to capture the type of Physical Specimen? or is it that the values (as John is suggesting) - that we are seeing in basisOfRecord - go in preparations?

Gary Motz (Indiana Univ.): SPNHC and others are evaluating the need for understanding how "Preparations" are used and whether or not it would be useful to propose as a full-fledged DwC extension.

10. Sophie Pamerlon: what about other DwC terms, is it planned in the (near) future to suggest controlled vocabularies for terms such as Habitat for example?

11. Jodi: I've been wondering if there's a way to mark a record, or better yet each field, as compliant with x version of a standard.

John: you can cheat, but there's no current way

Jodi: "Cheat", like prefacing the value with a code?

John Wieczorek: Like adding dynamicProperties to say the authority for each field.

12. Jodi: What about declaring your vocabulary(ies) so community/taxon-specific vocabularies can be used?

(Note/Question for others in DataMgmt IG) - Should we submit these questions to the DwC Hour Webform on behalf of these individuals? That way we can answer and document responses?

John Wieczorek: GBIF basisOfRecord dictionary: https://tinyurl.com/gkrlhb7
John Wieczorek: VertNet's lookup tables: https://github.com/tucotuco/DwCVocabs/tree/master/vocabs

Discussion

10 minute review https://github.com/tdwg/dwc-qa
Post-webinar survey: https://tinyurl.com/jxgk3uv/

Seems obvious we need a webinar on preparations!: Andy Bentley, University of Kansas Fish Curator, he's offered to give the webinar about preparations. //from Deb, he'd like to have assistance, a partner, to do this with.

Data Mgmt Interest Group - Meeting 21 Mar 2017

To Dos: Webinar presenters should ensure that their system checks are performing optimally (30 min prior to beginning of webinar)

- work on DMI events (next webinars, including DwC Hour, but others as well).
- Let's set up a regular meeting time. We can shoot for 1x month (and then skip a meeting if we don't need it).
    - Yes. We'll meet first Friday of the month, unless we need to change.

    - It should be a few days after the most recent DwC Hour - for best memory of events.
- Digest what we accomplished at last DwC Hour and *in general*, Review actions taken so far, Review Survey data feedback (if available), Darwin Core Hour Input Feedback
    - Follow-up activities - assignees
        - Gary offered to do the sound / share screen checks
        - Presenters asked to log in at least 30 minutes before the presentation to do a final sound check before starting.
    - Document this process for a DwC Hour Webinar and follow-up (on GitHub)
    - Choose next DwC Hour webinar topics / presenters
    - Dates / Times for these webinars
        - need a presentation for June!
    - Assignees to create Calendar Announcements, send out announcements to various WG / Lists
    - Who would
    - l like to moderate next time?
    - our collective thoughts...
- Merging “management” repo
    - use “manage” as label to sort / group?
- Process
    - show up for testing / hour before / day before
    - log in at least 30 minutes before webinar to test all is well
    - sound (please use headsets)
    - need volunteer to create mp4’s after Adobe Connect sessions
        - Gary Motz will check on his ability to do this...doesn’t have Camtasia
        - Sent email to Kevin to ask him to write-up the process.
- Access to repo (for editing)
- Moving issues from Google Doc to /dwc-qa repo

Current schedule:

NB: We should assign a “second” to moderate the presentation so that the “second” would be the predetermined responsible party for temporarily promoting question answer-ers to “Presenter” status and demoting them back to participant status

Related Item: Asking people to provide their affiliation with the Adobe Connect feature (see top right of share-pod for attendees).

April 3rd - Sound check for Paula (Gary - Moderator; John W. also present)
April 4th - “Thousands of shades of…” Paula Zermoglio (Presenter); Gary Motz and Deborah Paul (Moderators). Kevin will log in 30 minutes before to double-check sound.
May 9th - “The IPT and Darwin Core Archives” Joanna, Laura, Kyle (Presenters); TBD (Moderator)
June - (Presenters); TBD (Moderator)
July - “Georeferencing” (John W); TBD (Moderator)
August - (Presenters); TBD (Moderator)
Sept - (Presenters); TBD (Moderator)
Oct - (Presenters); TBD (Moderator)
Nov - (Presenters); TBD (Moderator)
Dec - (Presenters); TBD (Moderator)

Discussion Town Hall / Extraordinary DwC Hour

preparations (Andy Bentley and ?) // from Shelley - Randy S. has been working on some of this...
establishmentMeans (Quentin Groom - Belgium? and ?)
    Let's invite them to join forces - come to our next meeting?
    Send email first. Email sent 23 March 2017 DP
    They agreed, next steps (date / time / title etc, in progress).

aggregators perspectives ("added value / adding value to published data")
    iDigBio
    VertNet
    ALA
    GBIF

"Future DwC hours"
dwc Georeferencing Data - locality specific [date JULY 2017]
dwc The IPT and Darwin Core Archives [May, Kyle Braak, Laura Russell, Canadensys, Joanna, et al]
dwc Aggregator specific presentations (VertNet, iDigBio, ALA, GBIF, Canadensys, etc.)
    template talking points (send email to work on this). (material in tickets)? dwc:basisOfRecord
    joint
    earlier rather than later
    specific - brief rundown of workflows with data (10 minutes).
        - operations performed on data
    open discussion.
dwc Darwin Core and the Data Quality TDWG-GBIF group's work
dwc Kurator Workflows [John, et al] (September-ish)
    present when infrastructure at iDigBio (VM) is set up for web app
        likely to get very large hit when this happens.
dwc Let's Review DwC Tickets Hour and Issues so far / what's been asked / answered / and is still open
dwc More on Controlled Vocabularies - what's in a Taxon Name anyway?
dwc Geography - just where did you find that specimen?
dwc preparations (Andy Bentley and Quentin Groom) - being scheduled now with dwc:establishmentMeans
dwc and media
    audubon core
    3D data
dwc establishmentMeans (being scheduled with dwc:preparations)
dwc geography including waterBody (drainage, marine division, pelagic, ...) JW 2018 (paper!)
dwc capturing Stratigraphy, paleobiology, isotope data sharing, zooarchaeological data (JW, et al - Rob Guralnick, Laura Brenskelle, Kitty Emory (archaeology), Michelle LeFevbre (post doc of Kitty) - Send Note to invite - set date (2018)
dwc standards evolve, standard review planned by GBIF, how our input from dwc hour efforts help
dwc using the /data files to act on DQ efforts locally, "what's in your wallet?"
dwc challenges of combining paleo + neontological data in a dataset - dwc and ipt challenges (maybe ask Talia Karim to do this one)? - Send invite.

other "special guest" dwc hours - as needed and as time permits (to cover more topics faster than 1x month)
example: forming discipline-specific groups to work on improving the data in the fields where dwc suggests using a controlled vocabulary. Discuss how IT can only fix so much, then other communities (GBIF nodes, SPNHC, ...) must step in

Impressions from last Dwc #2
Participation +1
Support +1
helps us not to miss bits.

Darwin Core Hour 4 Apr 2017 (WEBINAR)- Notes

Presentation

Notes by Gary Motz, John Wieczorek, Shelley James, Deb Paul
Moderators: Gary Motz and Deb Paul. //from Deb - great job Gary!
(Adobe Connect reported 69 unique attendees)

Thousands of Shades for “Controlled” Vocabularies
Distinct values used by aggregators vs controlled vocabs (not the same)
Human data capture behaviour is an issue!
Language issues (between and within).
Capitalization (formats)
“Levels” (subsets)
“Modifiers” (contexts for behavior)
Random data that don’t appear to have meaning e.g. mismapping, collection conventions
Does order matter if there is a list of behaviors?
Is there a particular format for words and lists?
Taxon-dependent subsets?
Scope of an observation, one individual or many

Definition says it should be a description - is that conducive to a controlled vocabulary?

Conceptual issues - distinct words for one concept, one word for multiple concepts.

Issues due to: insufficient resources, participation, gaps in documentation
Solutions include: automation, best practice, community participation, documentation development and adherence, lookup tables

Lookup tables as a way to get from raw data to standardized data from controlled vocabularies.

Darwin Core standard distinct in scope (definitions and “official” recommendations) from Darwin Core Media Wiki (translations and “unofficial” discussion)

Words must represent a concept with single meaning for the community - agreement
Build controlled vocab using thesauri or ontologies - or both?
http://baskauf.blogspot.com.ar/2017/03/controlled-values-again.html
http://baskauf.blogspot.com.ar/2016/03/ontologies-thesauri-and-skos.html
Controlled vocabs do exist! Aggregators, providers, expert development (specific use e.g., Pensoft, Paris, AppleCore)
Legacy data - verbatim data - mapping to lookup tables

Discussion

Deb Paul (iDigBio, at FSU): Paula's presentation link will be up on the DwC Q&A Wiki: https://github.com/tdwg/dwc-qa/wiki after today's talk

John Wieczorek (Darwin Core): Recorded Presentation https://tinyurl.com/jcozm3l go to time offset 0:08:25

John Wieczorek (Darwin Core): That recording is about controlled vocabularies as well, with the issues around dwc:sex being the main example.

Deb Paul (iDigBio, at FSU): What's inside these dwc fields now? from GBIF, VertNet, iDigBio, ... Distinct Values seen in DwC fields suggesting use of controlled vocabs: https://github.com/tdwg/dwc-qa/tree/master/data

“Dead as a behavior”
Brad Millen: I can't understand how Dead on road is behaviour
Deb Paul (iDigBio, at FSU): Hm. Dead as a behavior...
John Wieczorek (Darwin Core): Nevermind crossing the road and the reasons for it.
Brad Millen: I is "Dead, Jim"
Deb Paul (iDigBio, at FSU): Not dead, resting, pinin' for the fjords :-)
Quentin Groom (Botanic Garden Meise, Belgium): reasons include the scope of an observation, one individual or many

DarwinCore Wiki and more resources (mostly from John, also Steve and Shelley)
Darwin Core Wiki for each Darwin Core term, where the URL follows the pattern http://terms.tdwg.org/wiki/dwc:[termname] For example, http://terms.tdwg.org/wiki/dwc:behavior
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: http://www.tdwg.org/standards/status-and-categories/
Category Data Standard: defined as "Specifies valid values in controlled vocabularies"

Steve's Baskauf blogs:--- Ontologies vs. thesauri ---
Mar 2017 (a bit technical) http://baskauf.blogspot.com.ar/2017/03/controlled-values-again.html
Mar 2016 (less technical) http://baskauf.blogspot.com.ar/2016/03/ontologies-thesauri-and-skos.html
http://rs.gbif.org/vocabulary/
Arctos http://arctos.database.museum/info/ctDocumentation.cfm
MNHN Paris http://standards-sinp.mnhn.fr/nomenclature/
VertNet Lookup tables: https://github.com/tucotuco/DwCVocabs
Apple Core: http://applecore.biowikifarm.net/wiki/Category:AppleCore_term
World Register of Marine Species: http://www.marinespecies.org/traits/wiki/Traits:Marine_species_traits
USGS Age: https://www.pwrc.usgs.gov/BBl/manual/age.cfm
--- Include vocabs in the standard? ---Steve's Baskauf blog Apr 2016
http://baskauf.blogspot.com.ar/2016/04/controlled-values-for-establishment.html
MIxS - Minimum Standards for any Sequence - http://gensc.org/mixs/
Environmental terms (biome, habitats, features within them) ENVO - https://bioportal.bioontology.org/ontologies/ENVO
Shelley: http://www.getty.edu/research/tools/vocabularies/tgn/

http://www.obofoundry.org/

Controlled vocabularies:
Viktor Senderov (Marie-Curie BIG4/ Pensoft): I've created a controlled vocabulary for taxonomic status terms based on actual usages in 4000 Pensoft articles.
John Wieczorek (Darwin Core): @Viktor. We'll want to see that. I think Paula will be asking who has what already.
Viktor Senderov (Marie-Curie BIG4/ Pensoft): Sure, I am planning to share it more widely by the end of the week, together with the OBKMS ontology and inference rules. :)
Pentcheff [NHMLA]: Tannenbaum — “The wonderful thing about standards [or controlled vocabular lists] is how many there are to choose from.”
Deb Paul (iDigBio, at FSU): I also suspect that sometimes there is a desire to capture the “verbatim” value for a concept - and would not want to lose this with having to choose from a controlled vocab.
Pentcheff [NHMLA]: Deb — yes. And as machine interpretation of natural language gets better and better, we will increasingly want to be able to refer back to the verbatim text. That’s contrary to my old-fashioned desire to atomize the crap out of everything.

Quentin Groom (Botanic Garden Meise, Belgium): In a few cases terms in Darwin Core have a field to explain the controlled vocabulary (e.g. geodeticDatum). In some cases couldn't we add a field to explain which vocabulary we are using?
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: From a management point of view, if there are many complicated controlled vocabularies, it would be a lot for the DwC group to manage. Communities of interest could do some of the heavy lifting
Jodi: Quentin, I think declaring your vocabulary is very important. But, at the dataset level, the record level, or the individual field level?
Quentin Groom (Botanic Garden Meise, Belgium): Jodi, given the way records have a life of their own you have to link the vocabulary specification in the observation
Pentcheff [NHMLA]: There’s an analogy here between a “standard” and (perhaps) what controlled vocabularies should be: it makes me think of an organization’s bylaws vs. their standing rules. Bylaws are nearly constitution-like — very hard to change, and that’s the way it should be. Standing Rules are (more) easily changeable in response to community demand.
Pentcheff [NHMLA]: From a social engineering perspective — yes. Vocabularies developed from within a research community are more likely to gain acceptance than those developed by well-meaning (and even smart) “outsiders”.
Jodi: Quentin, I would agree. At the least. I wonder if multiple vocabularies might be used in the same record, though. So would a field level declaration be needed?
Dan Stoner (iDigBio): It does not need to be all-or-none... in cases where a good external controlled vocab exists (like ISO Countries), use the external source. If it doesn't exist, the standard can still recommend the "best" known vocabulary, either externally (someone's github or web page), or... in the standard definition itself.

Pentcheff [NHMLA]: And, if adherence to the list is required, records with non-list values get automatically trashed... which may not be the best outcome.
John Wieczorek (Darwin Core): @Jodi In that case, another potential solution is to make a list from multiple sources and point to that.
John Wieczorek (Darwin Core): So far though, the lack of controlled vocabularies seems to be the bigger issue.
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: The solution to some of the issues listed here is handled by the fact that terms in TDWG standards (including controlled vocabularies) are identified by URIs. If a URI is

used, the source of the vocabulary is known. We have dwc: terms for verbatim (text) fields and dwciri: terms for storing the URIs.
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: All of the spelling variants are covered by properties of the term, whose URI identifier can be opaque. The properties of the term (alternative spellings) can be added to, or modified without invoking any standards process.
Pentcheff [NHMLA]: Is there a time-dependence to those URI references? Do we need to worry about whether the list has changed since the record was created? Or am I just getting way too far into the weeds? :)

Not all terms need the same approach!

Pentcheff [NHMLA]: It may be worthwhile to have a controlled vocabulary editor hierarchy. There are discipline-specific parts of vocabularies that should be edited by domain specialists. Then, above those editors, could be aggregating editors who build the recommended combined vocabulary for an entire Darwin Core term. This is similar to the editor hierarchy in WoRMS.
Dean Pentcheff [NHMLA]: Another way to "discover" some of the key community people who might be important to involve, is to look at who is contributing large numbers of records involving controlled vocabularies to aggregators.
Shelley:ontology aggregators e.g. obofoundry.org; archive and library communities
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: We are trying to re-solve problems that librarians have already solved http://www.niso.org/schemas/iso25964/ Unfortunately ISO 25964 is behind a paywall

GeologicContext
Jess Miller-Camp [UCR]: Do the various GeologicContext terms related to units of time have a controlled vocabulary? If they don't, the units recognized by ICS should be standard where they exist.
Gary Motz: Organizing geologic context using USGS information with consistent vocabulary: https://geo-nsdi.er.usgs.gov/talk/thesaurus/outline.html
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: See http://rs.tdwg.org/dwc/terms/guides/rdf/index.htm#3.6_dwciri:_terms_having_local_names_that_don%E2%80%99t_correspond_to which recommends ICS
Gary Motz: Geologic context terms thesaurus: https://www2.usgs.gov/science/about/
Steve Baskauf - Vanderbilt/TDWG VOMAG TG: Darwin Core RDF Guide, section 3.6 recommends ICS as a best practice
Amanda Millhouse (NMNH): Time-dependence is particularly an issue in geological contexts as well. Stratigraphic names may change in ranking (bed vs. member vs. formation) and the same is true with time periods as well...which is one reason why geologic context isn't as straightforward

//from Deb: we need to **make the process** of understanding how to help get engaged in controlled vocabularies - **transparent.**
//from Deb: I also suspect that sometimes there is a desire to capture the “verbatim” value for a concept - and would not want to lose this with having to choose from a controlled vocab.

//from Paula ”who built the controlled vocab” -- not always easy to determine, hinders use.

Shelley: Metadata for controlled vocabs - publish them as open resource
from Deb, What a great idea!!! Thanks Shelley - need to say that in the webinar! :-)

Data Mgmt Interest Group - Meeting 7 Apr 2017

Friday April 7th, next meeting - 11 AM EDT, 12 Noon ART
http://idigbio.adobeconnect.com/room

- Issues that need attention on /dwc-qa github

- Darwin Core Hour - Open Panel Feedback to Questions / Tickets from the Community. Have people participate.
- Issues Assignment
  responsible for document to then add "Answered" tag - okay for all of us to contribute answers
  not everything becomes a webinar
  we're distilling the output of the conversations and questions
  periodic review of the "answers" - before "Answered" tag is applied, and then, decide about adding to documentation
  Engage some more people...(yes!)

NEXT Webinar (May 2nd, 10am EST)

- Make sure we give an **Intro in the next webinar** to why preparations (John) establishmentMeans,occurrenceStatus are going to be presented together.
  People want to do more with the standard (the standard is working) - so the standard needs to evolve - get the standard to serve more purposes.
- Deb and John will moderate Preps and Ext webinar next month. 20mins each Quentin and Andy. Ask for ppts to be on google drive.

Future webinars

- Data Quality -- Kurator?
- Aggregators specific presentations - this is what we do to the data (in a row, 2 at a time?)... in one month...

https://github.com/tdwg/dwc-qa/issues

BLOG POST 1 for dwcHour
https://docs.google.com/document/d/1gS-4BlFoWqA72cWvn9dmLpXiHPSw9CleHVuzUMRkqWc/edit

- Blog post. Include what has been done, include potential new topics?, Should at a minimum publish the list of resources for "controlled" vocabularies contributed mostly by Paula, John, Steve, Viktor, Shelley, etc.

Darwin Core Hour 2 May 2017 (WEBINAR)- Notes

11AM EDT
Subject: dwc:preparations, dwc:occurrenceStatus and dwc:establishmentMeans

Andy and Quentin
(from Deb, for reference: known issues understanding these topics that could come up today. In general, those outside TDWG do not know how new terms come into being - nor do they know that standards go through a ratification process - community driven (at best). For extensions, they do not realize they can develop their own. Better practice is to develop them with the relevant community. And, these do not get ratified (do not get formally added to DwC). BUT, for those using the IPT to compile their datasets - only GBIF accepted extensions are present in registered IPT instances. New, non-GBIF accepted extensions are only usable in a TEST instance (non registered instance) of the IPT).

Gary: Intro to dwc hour logistics
John: intro to this dwc hour

Darwin Core is a living, evolving standard for biodiversity data, and is managed under TDWG.

To be useful, standards benefit from a certain degree of stability. But when that stability becomes an obstacle to usefulness, change is necessary.

From time to time, motivated groups led by people with energy and a distinct lack of fear challenge Darwin Core to do more.

Today's presentation illustrate aspects of the process to improve Darwin Core, using real and timely examples of Darwin Core terms that try to do too much, and in so doing, fall short of emerging real-world needs.

These are their stories...

PRESENTATION ANDY
Def. extension
DwC star schema, how extensions link with DwC standard.
Examples of extensions
(Question from Deb: how many here today, have preparations - for non-fish collections)?
Numbers in aggregators
Problems with preparations

- Multiple preparations, concatenations
- Concatenation various terms
- “Standard” issues. Variations in terms to name the same thing (eg EtOH, ethanol)
- Language variations and encoding issues
- Not preparation info (mismapping?)

→ Extension motivation, including controlled vocab
Extension fields -->important for research!, not for collection management.
Controlled vocabs fields, need discussion with community (particularly societies).
Possible issues, A. Collections practices, B. Data vocabs. C. aggregator functionality. Need community discussion.

COMMENTS
1. Encoding
Deb Paul - iDigBio: Encoding - we need UTF-8

John Wieczorek (Darwin Core): And only UTF-8, and throughout the lifetime of the data.
Brad Millen - Royal Ontario Museum, Toronto, Canada: I recently added some Japanese translations to my identificationRemarks field with the characters in a Unncode Font when exporting to UTF-8 they do not translate. why Anyone have an answer?
John Wieczorek (Darwin Core): @Brad Millen. Could they have been encoded in UTF-16?
Deb Paul - iDigBio: To Brad Millen - we can have a much longer conversation about UTF-8 and encoding issues. There are many places where in the sharing of data -the encoding can be garbled, or display improperly.
Brad Millen - Royal Ontario Museum, Toronto, Canada: I am not sure. A Japanese volunteer translated the old Yokohama Tading enclave labels. I never thought that.
Deb Paul - iDigBio: Brad, do you have a file they sent you with the translations? (Spreadsheet, Word doc? text file)? You can check the encoding
Brad Millen - Royal Ontario Museum, Toronto, Canada: I should have the original somewhere. UTF-16 never crossed my mind!

2. Suggested field:
Jess Miller-Camp (UCR): Might be worth it to have a field specifically for hazardous substances so people can just snap to that if that's what they're looking for.
Gary Motz (Indiana Univ.): http://manisnet.org/ASMdocstandards.pdf

3. Do researchers need to know where in a collection a preparation is stored?

4 . ***Patricia Mergen (Botanic Garden Meise/ Africamuseum):*** Many museums prefer not to share where the preparation are stored (to avoid robbery ...)

5. Logistics of creating an extension
Patricia Mergen (Botanic Garden Meise/ Africamuseum), speaking: Maybe the initial problem is that people don't understand how to use existing preparation field. Perhaps a controlled vocabulary could solve many of these issues. Also may be a good idea to look to other groups' standards before reinventing the wheel with this extension idea
Andy Bentley, speaking: Agree but beyond that the issue is multiple preparations, in particular users looking for tissue samples who cannot find this information buried in the preparations field.
Patricia, speaking*:* Have you checked with GGBN of existing standards they have determined already? Better to map between standards than start from scratch.

Question from John W. KEEP preservation history

6. ***David Shorthouse - CMN:*** Where and how do you wish to receive community feedback as extension is developed & when is the deadline for when extension will be submitted?
David Shorthouse - CMN: Then, when is the earliest it might be present as an extension in the IPT?
John Wieczorek (Darwin Core): @David There is no deadline right now. This is a bit of a first community discussion of the problem.
Deb Paul - iDigBio: David, I think folks will ask who would like to be involved in developing an extension for preparations.
Deb Paul - iDigBio: Also, David, no one has discussed this yet, but some work on the need for this (at least from the plant's point of view) has been done by the Apple Core group. Perhaps someone here today can explain the work of Apple Core

John Wieczorek (Darwin Core): @David I expect that a Darwin Core task group in TDWG may be created to follow through on this.
David Shorthouse - CMN: @John - would advocate for documents for groups who discover need for extension and want to get organized.
Deb Paul - iDigBio: @David @John - thanks - let's invite all here - please speak up if you'd like to be part of this preparations extension discussion / development effort

7. Discipline specific standards
Randy Singer, speaking: Very important to have discipline specific preparations controlled vocab, possibly even a drop down menu to force data providers' cooperation.
Randy Singer (iDigBio-FLMNH): Just to clarify on my point I'm just advocating for standardize, discipline vetted preparations to be used with another field for verbatim preparation. When you try to look across millions of data points it becomes almost impossible to get good data from non standardized prep types, but then verbatim prep type would exist for those that wanted the details and history of preparations. :)
Deb Paul - iDigBio: Yes @Randy - verbatim also good for Linked Data initiatives to link to things like BHL data

8. ***Carlos:*** How "preparations" of mounted specimens will be entered into that extension? There are specimens, either dry or wet preserved, that are mounted on something, like wood, plastic, glass, etc.

PRESENTATION QUENTIN

Observations vs. checklists. Different uses of DwC.
Where this is coming from (invasive species meetings).
Framework vs. standard. Framework is not a standard.
Questions around alien species. A. how did organism got there. B. where does it live. C. native? D. how well established.

a. How did organism got there. establishmentMeans. Def DwC. Suggested controlled vocab is a mix of stuff (eg invasive is how well established it is, not how it got there).
b. Where does it live. occurrenceStatus. Def DwC. Suggested controlled vocab again is a mix. (eg indications of abundance). How occurrenceStatus is currently used in GBIF.

Loose definitions make it easier to publish, stricter definitions make it more interoperable, but limit what can be published.
Vocab Maintenance Specification Task Group (TDWG)
GitHub repo with proposal:

a. Where does it live → New term dwc:origin. Def and proposed controlled vocab.
b. How well established → New term dwc:degreeOfEstablishment
c. How did it get here. → establishmentMeans, but change controlled vocab
d. Where does it live occurrenceStatus, but change controlled vocab

How you change DwC. TDWG
Questions to the community

COMMENTS
1. Absence data
Deb Paul - iDigBio: ooooh. Absence data!
David Shorthouse - CMN: Absence data with coordinates makes life as aggregator interesting.
Deb Paul - iDigBio: @David "we didn't see it - here :-)"
Teresa Mayfield (UTEP Biodiversity Collections): The absence data idea just blew me away....
Randy Singer (iDigBio-FLMNH): It could be in another database (e.g. occurrence database)

Randy Singer (iDigBio-FLMNH): the danger with "its not here data" could be that you just didn't see it/capture it though which is risky for research
Elizabeth Martin (USGS): But the metadata should mention whether part of the protocol included looking for a specific species. You will then know it was not seen even though the species was looked for.
Randy Singer (iDigBio-FLMNH): true you just need to be careful about comission errors and giving false positives (the "positive" being not invasive here or something like that)
Deb Paul - iDigBio: Yes @Randy - effort matters - where did you look exactly, how did you look / sample - how long did you look. But very useful if you go back to same spot repeatedly and want to keep a record of what is / isn't noted visit to visit. Hard to do well. Some are doing it well. Check with Andrew Short at KU - if you'd like to see an example of how this is being done from an entomologist point of view.
John Wieczorek (Darwin Core): See Darwin Core terms samplingProtocol (http://rs.tdwg.org/dwc/terms/index.htm#samplingProtocol) and samplingEffort (http://rs.tdwg.org/dwc/terms/index.htm#samplingEffort).
Deb Paul - iDigBio: @Randy - see CReAC - Collection Resources for Aquatic Coleoptera - a database that incorporates absence data. https://sites.google.com/site/theshortlab/creac
Randy Singer (iDigBio-FLMNH): thanks Deb

2. Verbatim vs. Formatted/Controlled Vocab
Patricia Mergen (Botanic Garden Meise/ Africamuseum): yes in ABCD you have for almost each concept an atomised field and the verbatim term (text field), it is very useful, but only very few providers take the time to fill it in nor to parse their verbatim data to put them in the atomised fields, unless it is like that in their database or they have an automated tool to split the data
Steve Baskauf (Vanderbilt University): Note that TDWG already has a way to differentiate between controlled vocabulary values and verbatim values. The IRIs that identify the controlled vocabulary terms are values of dwciri: terms. Verbatim values are values of dwc: terms.

3. ***David Shorthouse - CMN:*** Camels - dwc:origin = High Arctic.

4. ***Andy Bentley:*** Am I missing something or is this origin field more a taxon related field than a specimen related field?
Erica Krimmel, Chi. Acad. Sci.: @Andy I'm also curious about that
David Shorthouse - CMN: @Andy, that's my impression. But, what of fossil evidence?
Randy Singer (iDigBio-FLMNH): @Andy/Erica I would assume "origin" etc. would be subjective to the person doing the research. I'm not 100% sure how it would be helpful in the data, which are supposed to be unbiased
David Shorthouse - CMN: What of phylogenetic origin?
John Wieczorek (Darwin Core): @Andy Yes, it might be good to review a bit Darwin Core's capacity to share specimen data versus checklist data.
Beth Wommack: Could it be related to an individual organism if you found one in an area that was introduced? Say a specimen of an escaped falconry bird?
David Shorthouse - CMN: Or a zoo
Steve Baskauf (Vanderbilt University): Quentin and I have been experimenting with making a controlled vocabulary for occurrenceStatus conform to the machine-readable parts of the Standards Documentation Specification. You can read about it and see some demos at http://baskauf.blogspot.com/2017/05/using-tdwg-standards-documentation.html
Liz Sellers (USGS BISON): Zoo or Botanical Garden 'specimens' are often represented by basisOfRecord=LivingSpecimen
Gary Motz (Indiana Univ.): @Liz Unless the zoo specimen is contributed to a mammal collection after death.

Liz Sellers (USGS BISON): @Gary Yes indeed - would that then become basisOfRecord=PreservedSpecimen though right?
Gary Motz (Indiana Univ.): @Liz yes, indeed. But then, this could be where "origin" gets to be interesting!
Gary Motz (Indiana Univ.): How can we differentiate the "origin" of an African wild dog from the Toledo Zoo in Toledo, Ohio?
Teresa Mayfield (UTEP Biodiversity Collections): Or the African parrot received from a rescue facility.
Liz Sellers (USGS BISON): @Gary - Agreed. It will be interesting to say the least, to see just how something like 'Origin' (I also considered 'Provenance') will be implemented by those mapping their data to DwC. It's been 'fun' just trying to do this within the non-native/invasive species information management communities.

5. Interpretation in DwC data
Randy Singer (iDigBio-FLMNH): Is anyone else slightly worried about introducing interpretations to collections data? (e.g. this species is from X and it was established by Y) when this is interpreted data....can anyone convince me otherwise?
Patricia Mergen (Botanic Garden Meise/ Africamuseum): Well even an identification is interpreted data even if the name of the specimen is seen as part of the raw data …
Patricia Mergen (Botanic Garden Meise/ Africamuseum): Many of our users expect some sort of already interpreted data, but need metadata on how these interpretation was made
Randy Singer (iDigBio-FLMNH): @Patricia true but its basis is a means to locate a specimen within a collection, not an interpretation of the establishment means or invasive status of an organism
Quentin, speaking*:* If you have an individual that you see clearly planted or in captivity then this is clear. If you are observing a wild organism you may not have any basis to determine this, but if your observations evolve into a checklist then you may be able to determine the origin based on the research you are doing for your checklist. At the checklist level you may be talking about a taxon, at the occurrence level you may be talking about an individual organism.

6. On the process of changing DwC
Deb Paul - iDigBio: Sage Words from SPNHC: * SPNHC Strategic Plan 2013. “Because best practices are most effective when developed by, and applied to, the widest possible community, use the breadth of our membership as a tool to build collaboration with other stakeholders, including collection users and relevant professional societies (American Institute of Conservation, ICOM Conservation Committee, Natural Sciences Collections Association, AAM, discipline-specific societies).”

7. Scope of Quentin’s proposed changes to DwC
David Shorthouse - CMN: Quentin: Is scope of your proposal on organism (individual), community, or species?
David Shorthouse - CMN: ...or particular populations?

8. Occurrence vs. Checklist data
John Wieczorek (Darwin Core), speaking: Distinction between specimen data and occurrence checklists? Specifically about publishing dataset types (occurrence types vs. checklist types)
Quentin, speaking: Distinction between checklists and individual observations is not really clear in DwC, because it depends on the amount of time the collector/observer spends. Issues also with distinguishing between distinct individual organisms and organisms that occur as a group, e.g. clumps of plants.
John Wieczorek (Darwin Core): The suggestion was to discuss the difference between occurrence data sets and checklist data sets in Darwin Core archives.
Steve Baskauf (Vanderbilt University): We have a term in Darwin Core called dwc:organismScope to indicate whether the record is of a biological individual, herd, etc.

Quentin, speaking: @Steve yes, although there is not a controlled vocabulary for this term

9. At what level is does a preparation become a collection object?
Kevin Love - iDigBio: I have a question for Andy. When does a collections object of many preparations become many collection objects? For example, would an individual selected from a collection lot that is selected to be cleared and stained and the remaining individuals are preserved in ethanol... is this two collection objects? one collection object with two preparations?
Randy Singer (iDigBio-FLMNH): @Kevin/Andy Also wondering this
Teresa Mayfield (UTEP Biodiversity Collections): @Kevin We treat this as an object with two preparations.
Patricia Mergen (Botanic Garden Meise/ Africamuseum): @Kevin in ABCD we have a concept related or associated unit where you can handle to link one specimen from a lot uniquely to it
Patricia Mergen (Botanic Garden Meise/ Africamuseum): your can choses to have this within the lot record or interlink two or more records
Kevin Love - iDigBio: @Teresa... when might you treat them as two distinct objects?
Gary: How would this (proposed) extension clear up this confusion.
Andy, speaking: handled differently in different collections. If you are keeping the same catalog number then it remains one collection object, if an additional preparation is created and catalogued anew then this is a second collection object. This is a procedural difference between institutions. Doesn't really make much difference one way or the other. The preparations extension would be able to accommodate a 1:1 relationship just as easily as 1:many
Patricia Mergen (Botanic Garden Meise/ Africamuseum): Yes the tissue and DNA strains can be linked to the voucher ..
Randy Singer (iDigBio-FLMNH): separate databases for data associated with the same specimen makes informatics work really hard.
Brad Millen - Royal Ontario Museum, Toronto, Canada: The DWC2 field disposition come into play here for tissues
John, speaking: If you take a tissue sample out of a whole organism and then the tissue sample takes on a life of it's own this could be captured in DwC with the resource relationship extension. You can create subject-predicate-object, which is pretty wide open. Trick is to create a meaningful way to use these predicates, e.g. predicate "derived from." This predicate could also be much more specific, e.g. "DNA extract derived from"

10. Preparations for Paleo
Amanda Millhouse (NMNH Paleo): In regards to developing a preparations extension that could apply to paleo data, some data is only mapped to preparations because there is no other field in which to store it (eg vertebrate fossil morphology). So prep extensions to be used for paleo likely need to happen in conjunction with potential development of other fields that might be better for paleo data (or at least having an awareness of paleo data challenges)
Teresa Mayfield (UTEP Biodiversity Collections): @Amanda I am in agreement on the paleo issues.
Gary, speaking: One possible reason we see so many unique values in the preparation field is because people are using it as a catch-all a la Amanda's comments.
Andy, speaking: Agreed, and agreed that it is important to get representation from a broad span of disciplines.
John Wieczorek (Darwin Core): @Amanda and others. The Zooarchaeological community is also active in trying to figure out how best to capture collection objects ("Elements" in that community), which may have other interesting characteristics such as taphonomy.
Teresa Mayfield (UTEP Biodiversity Collections): @John is there a way for me to monitor that conversation? (Not sure I can contribute, but want to remain aware...)

Amanda Millhouse (NMNH Paleo): @John excellent! And yes, I'm happy to help give input for what types of paleo data are being mapped to dwc preparations
John Wieczorek (Darwin Core): @Teresa Yes, Kitty Emery, Rob Guralnick, Laura Brenskelle, and Michelle LeFebvre at UF are in the midst of a seed grant working on this topic.

11. People who have expressed a direct interest in being part of the on-going dwc:preparations discussion
    1. Amanda Millhouse (millhousea@si.edu) (paleo)
    2. Carlos Martínez (entomology) writes: I am interested in joining the discussion on preparations extension. Greifswald University. biotemail@gmail.com
    3. Beth Wommack ewommack@uwyo.edu (Vert paleo)
    4. Jess Miller-Camp (UCR Earth Sciences): If safety issues are something worth putting in preparation DWC, I'm happy to participate in that discussion. jessmc@ucr.edu Not sure if that's something to stay in house or not, though.
    5. Gary Motz garymotz@indiana.edu (representing zooarchaeology, paleontology, and botany...from a collection manager, biodiversity informatics manager, and paleo-researcher perspective)
    6. John Wieczorek gtuco.btuco@gmail.com (DwC, zooarchaeology)
    7. Andy Bentley (abentley@ku.edu) (fish)
    8. Randy Singer (rsinger@flmnh.ufl.edu) (fish)
    9. Deborah Paul (dpaul@fsu.edu) (botany)

Disciplines missing: herps, mammals, birds, botany

Many thanks to those who took these spectacular notes! Paula, Erica K.

Data Mgmt Interest Group - Meeting 5 May 2017

Friday May 5th, next meeting - 9:30 AM EDT, 10:30 AM ART
http://idigbio.adobeconnect.com/room/

Gary Motz: 64 total participants, 19 post-webinar survey respondents

Deb Paul: https://docs.google.com/document/d/1S_MzK5qhZVxGRNYDp6aslA2jZ-zo6m-KufLBqvc9IFY/edit#

Gary Motz:
https://apnews.com/7aefbd84420e4ab09540e21828de2374/Rhino-horn-stolen-from-University-of-Vermont;-reward-offered?

Gary Motz: I'd converted the AdobeConnect recording to MP4 and Kevin just uploaded to Vimeo.
Gary Motz: https://vimeo.com/216167534 for the Vimeo video.
Kevin Love: https://idigbio.adobeconnect.com/p46cvi3c2bi/ for AdobeConnect recording
Gary Motz: I don't have editing rights for the DwC-qa repo wiki page....could John or Deb post these links? Or grant editing permissions to me so that I can help out too?

Kevin Love:
http://clade.ansp.org/ichthyology/FTIP/search.php?mode=search&scope=Collection&contains=196140&contains_loc=&tbl=Specimens&Submit=Search+ANSP+Fish+Collection&gallery=ImageGallery

Kevin Love: http://search.idigbio.org/v2/meta/fields/records

Deb Paul:
https://zoologie.uni-greifswald.de/en/organization/departments/cytology-and-evolutionary-biology/staff/carlos-a-martinez-munoz/

Darwin Core Hour 13 Jun 2017 (WEBINAR)- Notes

Presenters: Kyle Braak - IPT Product Manager for GBIF, Laura Russell - Program Officer for Participation and Engagement at GBIF, Carole Sinou - Research Officer in Biodiversity for Canadensys.
Moderators: Gary Motz, Erica Krimmel

PRESENTATION

Kyle--*Into for newbies: what is the IPT & when should it/should it not be used*

- IPT Wiki: https://github.com/gbif/ipt/wiki
- IPT Website: http://www.gbif.org/ipt
- IPT article: https://doi.org/10.1371/journal.pone.0102623
- IPT example dataset: http://ipt.ala.org.au/resource?r=global
- and the same example, via DOI: http://doi.org/10.15468/qjgwba
- IPT data hosting centre: https://github.com/gbif/ipt/wiki/dataHostingCentres
- Sampling-event data: https://demo.gbif.org/sampling-event-data

Carole--*DwC-A and supported data sources in the IPT*

- GBIF Excel templates: http://www.gbif.org/newsroom/news/new-darwin-core-spreadsheet-templates
- Darwin Core Archives How-to-Guide: https://github.com/gbif/ipt/wiki/DwCAHowToGuide
- Darwin Core Type Vocabulary: http://rs.gbif.org/vocabulary/dwc/basis_of_record.xml
- Simple multimedia extension example: https://tools.gbif.org/dwca-validator/extension.do?id=gbif:Multimedia
- Audubon Media example: https://tools.gbif.org/dwca-validator/extension.do?id=http://rs.tdwg.org/ac/terms/Multimedia
- Identification History example: https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Identification
- Taxon core-vernacular names example: https://tools.gbif.org/dwca-validator/extension.do?id=gbif:VernacularName
- Event core-occurrence example: https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Occurrence
- Fossil specimen records mapped to “Occurrence” Core: http://rs.gbif.org/core/dwc_occurrence_2015-07-02.xml
- 1 or more image records for every fossil record mapped to “Audubon Media Description” Extension: http://rs.gbif.org/extension/ac/audubon.xml

Laura--*mapping source fields to DwC-A (live demo)*

> **Joanna McCaffrey (iDigBio):** also important to us is the dc:format field
>
> **Alex Thompson:** iDigBio also requires the dc:format field to be a valid MIME type (image/jpeg, movie/mp4, audio/mp3)

Kyle--*more for IPT admins; roadmap for IPT development*

- Ensure your IPT uses HTTPS and is always kept up to date! The latest version (v2.3.4) patches a security vulnerability: https://lists.gbif.org/pipermail/ipt/2017-March/000671.html
- Essential criteria for IPT data hosting centres: https://github.com/gbif/ipt/wiki/dataHostingCentres#data-hosting-centre-criteria
- Consider demonstrating to your users and funders that your IPT is sustainable and trustworthy by becoming a certified data repository: https://github.com/gbif/ipt/wiki/dataHostingCentres#certification
- Future IPT enhancement-making resource creation easier: https://github.com/gbif/ipt/issues/1198 and https://github.com/gbif/ipt/issues/1133
- Future IPT enhancement-Allow resource administration notes to be kept: https://github.com/gbif/ipt/issues/791
    - **Brad Millen, ROM, Toronto, Canada:** Great Feature
- Future IPT enhancement- Simplifying contact entry in resource metadata: https://github.com/gbif/ipt/issues/1166
- Future IPT enhancement-Enable dataset peer review: https://github.com/gbif/ipt/issues/1318
- Future IPT enhancement-Make the tool easier to install and update (e.g. CentOS packages): https://github.com/gbif/ipt/issues/1304
-

COMMENTS

1. IPT versioning

Liz Sellers (USGS BISON, Reston VA): Does it really "automatically" keep up-to-date with the latest DwC and metadata standard versions? Or only when you install/upgrade to the most recent 'version' of the IPT?

Laura Russell (GBIFS): @Liz, the admin syncs all cores and extensions. When a new DwC term gets added, the related cores and extensions get updated and new versions of those released. The IPT syncs with those, and automatically updates all existing dataset's mappings to use the latest DwC terms. More info about how t this works can be found in the IPT user manual .

Laura Russell (GBIFS), speaking: When an update is required you'll see a message on the admin portal of the IPT.

Liz Sellers (USGS BISON, Reston VA): @Laura Russell, Cool. Thanks. I was indeed not aware of that capability in the IPT.

2. DOIs on GBIF

Elizabeth Martin (USGS): I see the DOI is shown in the GBIF portal when the dataset was indexed. Does that DOI somehow is linked back to the IPT record that was generated in an installation of the IPT at a center?

Kyle, speaking: If your dataset has already been assigned a DOI then GBIF will not reassign it a new DOI. If it does not already have one, GBIF will assign one. The DOI on the IPT resolves to the GBIF page, which links back to the IPT.

3. OccurrenceIDs

Aaron Goldberg (UW-Madison): Can the GBIF IPT assign an occurence ID if they are missing for a dataset, or should those be assigned before attempting to pubish?

Kyle, speaking: They need to be assigned before you attempt to publish, and the occurrenceID should be globally unique. Best practice is not to use a sequential number assigned by your collection; instead, assign an occurrence ID that can last the lifespan of the record.

Erica Krimmel (Chicago Acad. Sci.): Any recommendations on where to mint unique occurrenceIDs? (if you are not using a CMS that helps with that)

Kyle, speaking: Probably most common method is to use a combination of namespace + number. As long as it's unique and persistent.

Aaron Goldberg (UW-Madison): Excellent, thanks.

Joanna McCaffrey (iDigBio): there are native UUID fields in some databases

Joanna McCaffrey (iDigBio): we just had instance where they changed their collectionCode, and they were using it in their triplet, so it made a mess.

4. Summary data
James (CORDIO EA): Can GBIF IPT present summary data/information on species occurrence etc?

Kyle, speaking: IPT cannot, but GBIF portal can.

Gary Motz (Indiana Univ): @James, there are a number of data summaries provided by a number of different data aggregators (including GBIF) posted on our GitHub here: https://github.com/tdwg/dwc-qa/tree/master/data

Joanna McCaffrey (iDigBio): tell them about the relationship extension if they change identifiers.

Laura, speaking: While it's preferable to always keep the same occurrenceID, sometimes it's not possible. (e.g. moving from Arctos to Specify). In this case, you can use the resource relationship extension to accomodate this by setting what the old and new occurrenceIDs are. This allows GBIF to update the occurrenceID, and also stays tied to the record.

5. Where does data published on the IPT go?
Deb Lewis: Are datasets published to GBIF "harvested" by iDigBio and vice-versa?

Laura, speaking: With VertNet, we'd publish the information on an institution's behalf via the IPT, so the data would end up going to both places. iDigBio has done some work to harvest IPT-available data from other institutions. We've always tried to make sure datasets are in both iDigBio and GBIF, but not everything has been.

Joanna McCaffrey (iDigBio): sorry, I had myself muted! The implicated of harvesting from GBIF might say we dont' ask, but we always ask for permission.

Laura/Kyle, speaking: rule is to always check with data providers before publishing to GBIF. Accept data sharing agreement, etc. to prevent data from accidentally being shared when the provider does not want this.

Laura, speaking: Not everyone is using the IPT. E.g. Symbiota uses the GBIF API but not the IPT, so there are some variations in how people publish.

6. Publishing vs. registering

Gary Motz (Indiana Univ): Could you clarify the difference or relationship between "publishing" a dataset and "registering" a dataset with GBIF?

Laura, speaking: Publishing is creating the DwC-A. Registering makes this DwC-A file discoverable through the GBIF registry API. This also adds the data to the GBIF portal.

Carole, speaking:

Kyle Braak GBIFS: What is data publishing?In the context of GBIF, 'publishing biodiversity data' is the process through which biodiversity datasets are made publicly accessible in a standardized format, via an online access point. Publishing the data following international standards enables integration of the data into GBIF.org where published datasets can then be discovered and accessed via the global GBIF.org website and associated web services thereby making it easily accessible by taxon, region, time period and other criteria.

7. Endorsement

Laura, speaking: USGS endorses the new datasets for U.S. collections. This is part of the GBIF endorsing process. For other countries information goes to the country's national node will endorse, or else if there is no country node VertNet/endorsing committee will review your data to ensure that it's appropriate for publishing on GBIF.

- Info about endorsement: http://www.gbif.org/publishing-data/endorsement
- Request endorsement form: http://www.gbif.org/publishing-data/request-endorsement#/intro

Kyle Braak GBIFS: To quote the form: "The process seeks to confirm the quality, interoperability and discoverability of data published through GBIF and to ensure that publishers receive proper credit and attribution"

Darwin Core Hour 11 Jul 2017 (WEBINAR)- Notes

Georeference Terms in Darwin Core: Where am I exactly? Presented by John Wieczorek and David Bloom

> //request from Deb, please, perhaps in a different dwc hour - something like "research use of nhc georeferenced data: implications from dwc term expectations/implementation and research expectations/implementation" and cover known issues, like datum (known / unknown), use of uncertaintyRadius or footprintWKT if provided, what "can" be done, what "could" be done if 1) researchers only had "x" and 2) what would be great for researchers collecting new data, to do in the future. As in the relationship of DQ provided by researchers to the downstream research possible / not possible (best practices).
>
> For this scheduled #dwcHour, the scope is to discuss / elaborate on #locality terms in dwc to explicitly talk about best practices for what should go in them / expect to find. And I assume / hope, that there'll be examples from our /data pile about what we see now, and at least a small discussion of ways in which we can all make it better going forward (IT / researcher / coll mgr - data mgr / citSci).

PRESENTATION

- Differences between coordinatePrecision and coordinateUncertaintyInMeters, both related to georeferencing concepts.
- Class/Property/Value in DwC
- Location terms in DwC: Geography - Georeference
- In the collection: specimen with and without coordinates
- Specimen with coordinates:
    Locality expressed with coordinates in deg-min
    Verbatim data and interpretations → verbatimCoordinates, coordinatePrecision, verbatimLocality, inferring coordinatePrecision from coordinates given.
    coordinatePrecision to nearest degree/half degree/tenth degree → location is not preserved by reducing precision.
        Effect of precision on Scale
        GPS limits: setting gives different spatial
    coordinatePrecision by latitude, precision does not give distance indicator, but a latitude degrees indicator
    coordinateUncertaintyInMeters, DwC def
    coordinateUncertaintyInMeters from coordinatePrecision, depends on latitude considered. Refer to point-radius method.
    Sources of coord uncertainty
- Specimen without coordinates:
    Use georef calculator, with coordinates from GMaps
- Precision vs uncertainty, vary according to the source
- Using records with coordinates.
- combining with environmental data
- What can coordinates say about precision and uncertainty: nothing.

COMMENTS

1. RECORDING & INTERPRETING VERBATIM LOCALITY INFORMATION

David Shorthouse (Canadian Museum of Nature): Would you leave verbatim locality blank in this case? [referring to verbatimLocality being coordinates in deg/min on the first example specimen card shown in the webinar presentation]

David Bloom: @David - can we answer your question at the end? There are a couple ways that folks can handle this field.

David Bloom, speaking: My take is that if you have the information, best practice is to fill it in to avoid possibly losing the data in the future. You could think of verbatim locality as the absolute and decimal lat/lon as the current best interpretation of the absolute information.

John Wieczorek, speaking: Agreed.

Peter: Is it fair to interpret the locality by knowing that the specimen was collected from a trout and therefore the locality must be a stream in the area

John Wieczorek, speaking: Yes, but depends on georef process (georeference several localities at once vs do it taking taxonomy into account, eg, fish would be on streams, it's an interpretation. Process must be described in georeferenceRemarks.

cindy opitz (university of iowa MNH): or a trout farm, fish hatchery?

Peter: Cindy, true enough.... :>)

David Shorthouse (Canadian Museum of Nature): In this case, it could have been swallowed several km upstream/downstream.
John Wieczorek: Another example: zooarch, spp carried around by people before actual location.
David Shorthouse (Canadian Museum of Nature): Like camels in the arctic.
Dave Bloom, speaking: sources of uncertainty: the more you report the better, good descriptions can make a big difference, give end users more info and knowledge and allow them to know if record is fit for their use.
John Wieczorek, speaking: sources of uncertainty play different roles according to the place, so better to report as many as possible.

Jean W (DMNH): What do you suggest for older specimens which just give a name, however you suspect they also collected outside of town?
John Wieczorek, speaking: refer to Quick Reference Guide: named place only. Half distance between nearest next named place.
Deb Paul: Georeferencing Quick Reference Guide: Georeferencing Quick Reference Guide: http://manisnet.org/GeoreferencingQuickReferenceGuide.pdf

2. RECORDING PRECISION
Deb Paul: So precision comes from the GPS?
David Bloom: Good question Deb. We'll answer when John wraps up.
John Wieczorek, speaking: Precision comes from GPS based on what the GPS screen is showing (i.e. what mode the device is in). So if the device is in degrees-minutes-seconds mode, then the GPS will give you coordinates with precision of the nearest tenth of a second. Precision is a constant with the GPS device, whereas uncertainty is very much conditional.
Deb Paul, speaking: So, for future best practice would it be best that the data provider look at their GPS unit and report the precision? Rather than for data managers to determine the precision retroactively.
John Wieczorek, speaking: Yes, although an easier way for the data provider to capture the necessary info is to just record the GPS device's accuracy.

David Shorthouse (Canadian Museum of Nature): Not-so-funny story about coordinates on IP addresses: http://bit.ly/2tJzlJq

> dp to ds: oh my goodness - not nice at all.

DP: Please add your Georeferencing Protocols and Workflows here https://www.idigbio.org/wiki/index.php/Georeferencing#Georeferencing_Community_Protocols_and_Workflows

DP: Please note, for best practices on what locality data to capture in-the-field, there's a guide you can use as a collector, or share with collectors providing you with specimens and related data - for your collections and for future research use!
http://www.idigbio.org/sites/default/files/working-groups/gwg/GoodBadLocalitiesV27Oct2015.doc

PZ: Also see: MVZ Guide for Recording Localities in Field Notes:
http://mvz.berkeley.edu/Locality_Field_Recording_Notebooks.html

> Hey Paula, FYI: the above version at iDigBio is an updated version (2015) of this one that you reference (2005). It includes the same information - but adds references about Darwin Core terms for this locality data collected and it includes recommendations about what to do if you're using phone gps apps.

DP: help stop the proliferation of "legacy" data :-) get data "born digital"

Data Mgmt Interest Group - Meeting 14 Jul 2017

Present: John, Erica, Gary, Paula, Deb

DwC Debrief 14 July 2017.

Effort - John's point about how much time effort takes to put this together.
Gary Motz - putting self out there as interested / vested party - presenting the issue.
John - Risk Mgmt - we need more people - to spread the workload.

> Invites sent (by dp) to: Shannon Ascencio, David Shorthouse, Dean Pentcheff, Holly Little, Arthur Chapman (time zone issues) (Data Quality), Sharon Grant, Peter Desmet, Dag Endresen, Rick Levy

Laura Brenskelle? - presentation - zooarchaeological work (two parts, one prob in common with paleo - trying to accommodate deep time - work with paleo person to put this together)

TDWG (not DwC Hour) [Walter Berendsohn (TDWG origins were around controlled vocabularies)... historical perspective.] ABCD development. TDWG Steve Baskauf

How are you using DwC, What you are challenged by using DwC Core? ... 5 - 7 minute talks?

Latin America outreach region (Colombia, Ecuador, Argentina, Brazil - Antonio Saraiva) CONABIO, Renato De Giovanni (CRIA, Tapir)

Review the GitHub tickets
At some point - get together to review GitHub Issues Tracker - when is good? **WhenIsGood sent 16 August (dp)**
Last two weeks of August - when is good - review github tickets - issue tracker

Darwin Core Hour Plenary Talk at #TDWG2017 - planning

dwc hour idea - shared with John on January 5th 2017

Darwin Core Hour 15 Aug 2017 (WEBINAR)- Notes

Title: Darwin Core Hour: GBIF and iDigBio present the Aggregator's Viewpoint
Part ONE: GBIF: free and open access to global biodiversity data
Part TWO: iDigBio - aggregating and enhancing vouchered global biocollections data

Abstract. In this next Darwin Core Hour Series, we shift to the viewpoint of large biodiversity data aggregators. In this session, we welcome GBIF and iDigBio. GBIF aggregates the world's biodiversity data from observations to checklists to biological specimen data. iDigBio aggregates vouchered specimen data with related media, genetic information and trait data. The Darwin Core Standard plays a key role in the standardization of biodiversity data and in the design and implementation of strategies to improve data quality. Many people wonder what happens to their data after they provide it to an aggregator. Find out the answers to such questions as: what does the aggregator do to assess fitness of the data?, what are the most common

data issues seen?, what does the aggregator do to data to make it easier to find when searching an aggregator database?, and how does sharing data with an aggregator benefit me as a collection manager/curator/researcher/data scientist?

PRESENTATION

1. **GBIF (Kyle)**

- Intro to GBIF, history of it.
- Data discovery
- presenting gbif.demo.org. Invitation to test it and provide feedback
- data assessment and preparation. Classes of data, indexing, endorsement, automated checks
- message: first publish
- workflow to get data into GBIF.
- indexing workflow, ingestion workflow. Detection of issues and flagging. Raw, verbatim and interpreted data.
- value added: what does gbif do to make data searchable.
- Data metrics / data statistics
- Updating data.

(Andrea)

- Aggregator outreach, working with other aggregators.
- Aggregator future, next steps
- Weak points and biggest challenges
- Strong points
- Value of stakeholders

2. **iDigBio (Deb)**

- intro / history iDigBio, differences w/ GBIF: only preserved specimen records
- data discovery. Challenges in the datasets (eg occID, mapping issues)
- data assessment and preparation
- required / recommended terms to include.
- data issues seen.
- ingestion workflow. Ingestion and indexing.
- value added. Metrics, flags, maps, registration with GBIF. Data services.
- aggregator outreach.
- future.
- weak points: future funding, design of the portal (aggregation), standards have limitations.
- biggest hurdle: data quality
- strong points: cyberinfrastructure, data quality flags, IPT and assistance, community building, research initiatives.
- value for stakeholders, table with benefits for each kind of stakeholder.

COMMENTS

Questions pre-webinar:

1. DATA QUALITY REPORTING MECHANISMS

Brad Millen, ROM, Toronto, Canada: I know that GBIF has a mechanism for users to report Data Issues. Does GBIF keep track of these in any way like VertNet does in GitHub? Does iDigBio? If iDigBio has a reporting mechanism how may I access it? and if not does iDigBio plan on enabling this tool?
If a user asks of a query of the aggregated data and see a record that prompts a question about the veracity of that data.

Brad Millen, ROM, Toronto, Canada: Perhaps at end of webinar this can be answered

Joanna (iDigBio): Brad - please say more about 'data issues'.

Brad Millen, ROM, Toronto, Canada: If a user asks of a query of the aggregated data and see a record that prompts a question about the veracity of that data

Joanna (iDigBio): Brad - the current mechanism is for the finder to report issues to the provider

Brad Millen, ROM, Toronto, Canada: OK

Kyle (talking): GBIF has a feedback mechanism built in, sent to the publisher via GitHub and categorized by type, e.g. bug or data qualoty

Brad Millen, ROM, Toronto, Canada: Part of my question(s) posted at beginning of webinar in Chat was partially answered by Kyle. The rest …….. Can this be answered.

Alex Thompson: iDigBio has a feedback reporting for that creates internal tickets for us

Joann (talking): we don't have any explicit place for iDigBio to collect comments on the data

Deb (talking): we don't yet have an automatic system for tracking feedback but we do go through and manually create GitHub tickets

Alex Thompson: Often that involves us relaying information back and forth to the provider when issues are discovered

Brad Millen, ROM, Toronto, Canada: Understood. Many thanks.

Webinar started GBIF QUESTIONS.

Brad Millen, ROM, Toronto, Canada: 797,576,089 occurrences right now - referring to GBIF

Deb Paul: Yep - that's occurrences as in observations data, sample data, literature records, and specimen records data

John Wieczorek: GBIF Demo Portal at https://demo.gbif.org/

2. DIFFERENCE BETWEEN OCCURRENCE AND SAMPLING EVENT?
Annie Simpson: what is the difference between occurrence and sampling event? IAre sampling events always repeat data on previously entered taxa?

John Wieczorek: I'll try to answer the Occurrence versus Event differences here.

John Wieczorek: Occurrence and Sampling Event are two perspectives on species at a place and time. The distinction puts the focus on organisms in the Occurrence case and on the time/place/protocol in the Sampling Event case. Both can produce Darwin Core Occurrences, but because of the limitations on structure in an Darwin Core Archive, Events might need to be the core record. This is to be able to connect extensions with, for examples, measurements on the Events as opposed to on the organisms.

John Wieczorek: The limitation is the star schema.

Annie Simpson: That is fine for now; it is much more complex than I had initially thought however.

3. DIFFERENCE BETWEEN RAW AND VERBATIM DATA?
Annie Simpson: What is the difference between raw data and verbatim data?
Kyle (talking): store original xml responses. You may have 1 xml response, raw response, that has verbatim (after processing issues you get interpreted data).

4. TRACKING DATA CITED IN PUBLICATIONS
Jen Hammock: How are data record mentions in publications detected?

Kyle (talking): feature available on demo.gbif.org, but GBIF has been working to better integrate citations within published datasets.

Jen Hammock: DOIs, I think?

Jen Hammock: ohhhhh

5. HOW CAN I COMMENT ON A DATASET?
Jean Woods (Delaware NMH): can someone post the link for comment on GBI

Jean Woods (Delaware NMH): Can someone post the link for commenting on GBIF usage statistics? Thanks,

Deb Paul: GBIF Call for Input on Dataset Metrics: Call for input: https://github.com/gbif/portal16/issues/138

John Wieczorek: @Jean Woods Have a look at this page for your birds collection to see citations and statistics on the data set. https://demo.gbif.org/dataset/c21cd435-718a-4069-b503-776bf0e22b96

Deb Paul: @Jean Woods - you got the link?

John Wieczorek: At the top of that dataset page is an icon for feedback.

6. SUMMARIZING DATA USE FOR DATA PROVIDERS
Brad Millen, ROM, Toronto, Canada: Interesting!

7. FORMAT OF DOWNLOADED DATA

Paula Zermoglio (University of Buenos Aires): Kyle, Andrea, when you download data from gbif after a particular search, what do you get? raw, verbatim or interpreted data? or all of them? if some, how can we access the rest ? from the DwC-A?
Andrea (talking): if you download the DwC-A you get verbatim and interpreted in different tables.

Brad Millen, ROM, Toronto, Canada: Great question. Beat me to it.

Deb Paul: @Paula - all three but not sure of format of download. @Kyle?

8. FORMAT FOR PUBLISHING COLLECTION INVENTORIES?
Wouter Addink: is there a way to share with GBIF inventories of collections on storage level (lists of drawers and jars with location and some metadata what they contain (e.g. on species, genus or sometimes higher level)

Erica Krimmel: @Wouter you may be able to publish that level of info via a GBIF 'checklist'

Wouter Addink (Naturalis): you mean I can use a storage unit as basis for a checklist rather than a species?

Erica Krimmel: Best to put this to Kyle and Andrea, but I think a checklist generally includes taxonomy but can be themed around non-taxonomic characters..

John Wieczorek: @Wouter This is interesting, and not the first time I have heard this recently. The question each time has been, is this information that collections actually want to share?

Wouter Addink (Naturalis): I am thinking of DiSSCo here where we want to create an inventory of all collections in Europe as first step of further digitisation

John Wieczorek: I see. So sharing information about the scope of the pending digitization challenge.

John Wieczorek: Sounds good for an extension.

Wouter Addink (Naturalis): so we have an idea what is in the collections and where it is.

Andrea (speaking): might also be a good case for a sampling dataset, because the inventory could be considered an atypical kind of event. Taxonomy can be at any level, i.e. “a box of beetles” is good enough.

Wouter Addink (Naturalis): Naturalis has an example dataset, which is currently shared as DwC (which I think is not appropriate for this)

Alex Thompson: @Wouter This is a good potential use case for NCD

Wouter Addink (Naturalis): https://demo.gbif.org/dataset/62d82928-dc6f-40dc-85b3-f2be47e7b49a

Wouter Addink (Naturalis): @Alex yes, although I see NCD more as a description on collection level

Gary Motz: NCD - Natural Collections Description - http://bioimages.vanderbilt.edu/pages/NCD-v090_TDWG-NonNormative.pdf

Alex Thompson: We will be working on the new one this year at TDWG

Alex Thompson: Watch: https://github.com/tdwg/ncd for development

Gary Motz: I was attempting to decipher the acronym for those that didn't know what it was (myself included). Thanks for the additional info.

John Wieczorek: @Wouter I see the data set. I can see how this would be a pain to do via metadata-only resources.

John Wieczorek: @Wouter Does ABCD have any support for the collection-level storage information?

Wouter Addink (Naturalis): @John, I have not looked into that yet. Might be.

Brad Millen, ROM, Toronto, Canada: Polling our records from ROM IPT of Vertebrates

9. DATA HOSTING BY GBIF
Liz Sellers (USGS BISON, Reston VA): What's the ETA on the data hosting service? Months? Years?

Andrea Hahn (GBIFS): @Liz: if your question relates to central hosting services, then it is available already to some degree for specific sections of our publishers,

Andrea Hahn (GBIFS): for example through a cloud-based IPT available to publishers within the BID (Biodiversity Information for Development)

Andrea Hahn (GBIFS): via a cloud-based IPT installation

iDigBio QUESTIONS.

1. MEDIA ON IDIGBIO
Gary Motz: @Deb, @Joanna, @Alex, or @Matt, for 3D meshes, is that specifically the *.mesh file type or the class of data that is "meshes" broadly speaking

Alex Thompson: @Gary Motz: Right now we use

Alex Thompson: STL files

Alex Thompson: but we only have a handful of providers

Gary Motz: Doug Boyer and I are working on developing metadata recommendations for 3D objects so that 3D data provision barriers can be lessened. Stay tuned for more information on that in our DwC Hour presentation in November

2. GEOGRAPHY DATA CLEANUP/ISSUES
John Wieczorek: 197 countries, 250 country/territory country codes

Brad Millen, ROM, Toronto, Canada: Georeference output may be different than original verbatimLocality. Distinction sometimes

Alex Thompson: @Brad we support the full range of darwin core data fields, and try our best to preserve and present verbatim data so that sources of errors can be traced

Deb Paul (talking): state = Oregon, so why is there a georeferenced point off the coast of Africa??

Jodi: 0,0

Deb Paul (talking): right! Whoever sent us this data had entered zeroes as lat/lon placeholders rather than leaving them null. Don't do this.

Alex Thompson: @Deb or a database column has an inappropriate default value that got autofilled

3. IDIGBIO SERVICES (ESP. FOR NON-US INSTITUTIONS AND NON-US SPECIMENS)
Alex Thompson: iDigBio also has cloud hosting services if you want to run your own IPT but don't have a server to do it on

Gary Motz: https://www.idigbio.org/wiki/index.php/IDigBio_API#Client_Libraries

Alex Thompson: in addition to running an IPT server on which we will host your datasets

Brad Millen, ROM, Toronto, Canada: Even for those outside USA

Joanna (iDigBio): We can't register outside-US datasets with GBIF

Alex Thompson: But we can still help you host your IPT

Brad Millen, ROM, Toronto, Canada: Ok, I wondered.

Joanna (iDigBio): @Brad http://ipt.idigbio.org

Brad Millen, ROM, Toronto, Canada: ROM http://gbif.rom.on.ca/ipt/

Brad Millen, ROM, Toronto, Canada: Tweaking it currently

Paula Zermoglio (University of Buenos Aires): Deb, would iDigBio deal with datasets from non-US countries? or is it mainly US-centered?

Matthew Collins: https://jupyter.idigbio.org

Joanna (iDigBio): @Paula - we are interested in any specimen data *[seconded by Deb, so long as they are vouchered]*

Paula Zermoglio (University of Buenos Aires): thanks joanna and deb

Alex Thompson: GUODA Shoutout to Jen Hammock (EOL)

4. WHERE TO PUBLISH DATA?

Wouter Addink (Naturalis): should data providers provide their dataset to as many aggregators as possible in order to benefit from the tools building on specific aggregators?

Wouter Addink (Naturalis): or do we have a huge data duplication problem then

Deb (talking): Yes. Different aggregators focus on different data users and can provide different data quality/feedback mechanisms. Discoverability is a huge issue, so that takes precedence over a fear of duplicate data being published across aggregators.

John Wieczorek: I second that, Deb.

Andrea Hahn (GBIFS): Yep. We will have to handle duplication on all levels anyway

Andrea Hahn (GBIFS): Try to keep record identifiers as stable as possible though

John Wieczorek: Until one aggregator can enable the community to make custom views and indexes anyway.

Alex Thompson: @Wouter Don Hobern and I have talked about working on an infrastructure to solve the duplication (and duplication of effort) problem

Wouter Addink (Naturalis): will aggregators aggregate the data from each other then

Alex Thompson: but we haven't gotten anywhere yet

Brad Millen, ROM, Toronto, Canada: Canadensys is essentially Inverts, Plants etc. but they come to ROM IPT and poll records and publish on their site.

cindy opitz, university of iowa MNH: but don't the big aggregators pick up data from others? do we have to submit data to multiple aggregators?

Alex Thompson: all of the aggregators are directly sourced, we don't exchange data internally

Joanna (iDigBio): @cindy - we try very hard not to intentionally ingest duplicates.

Matthew Collins: If you make your data available (via IPT or other methods) then the relationship is more that multiple providers can get data from you.

Alex Thompson: we all build on the same infrastructures though, IPT, DWCA, etc.

Gary (speaking): if publishing in several portals, issues identified in diff portals would go back to the provider, responsibility falls on the provider.

Deb (speaking): In some portals updates might be faster, then one portal could have a more updated dataset. Versions, with ids associated. Data is sitting in IPT, then all aggregators can grab from the same place. Eg: GBIF was already there and publishing when iDigBio came to be, they would not grab datasets they already have, but what's new through idigbio.Need for robust identifiers to avoid duplicates.

Andrea (speaking): aggregators are not ingesting everybody else's data.

John Wieczorek: @Cindy One way to think of it is that our data sets on IPTs are like market places. Aggregators come to get the data of interest.

cindy opitz, university of iowa MNH: I'm confused. I submitted data to VertNet, and now it is also in iDigBio and GBIF

Brad Millen, ROM, Toronto, Canada: Records from ROM are appearing from our ROM IPT and/or GBIF on many sites.

Alex Thompson: Its useful to think about vertnet as two parts: A data publishing service (ipt.vertnet.org), and an aggregator (vertnet portal)

John Wieczorek: @Cindy You publish data. VertNet facilitates making sure aggregators get it.

cindy opitz, university of iowa MNH: yes?

cindy opitz, university of iowa MNH: ah, thanks

Brad Millen, ROM, Toronto, Canada: Once published anywhere it spreads rapidly.

Alex Thompson: @Wouter think of it less as exchanging data, and more of a global harvesting layer built on top of IPT that is then available for other people to build portals off of

Alex Thompson: like John was talking about with custom views and indexes

Joanna (iDigBio):
https://www.idigbio.org/wiki/index.php/Data_Ingestion_Guidance#Instructions_on_changing_identifiers_.28occurrenceID.29

John Wieczorek: This identifier change issue is important and a bit complex. VertNet helps people get through this and build a ResourceRelationship extension for the data set. GBIF and iDigBio both use that to make the change.

Wouter Addink (Naturalis): A central authority is needed to create the persistent identifiers

John Wieczorek: GBIF gets them from VertNet. (speaking of ResourceRelationship extensions mapping old occurrenceIDs to new ones)

Data Mgmt Interest Group - Meeting 18 Aug 2017

9 am EDT

Welcome any newcomers - results of outreach for new members to DCH Team
    answer their questions / share our process documentation
        https://github.com/VertNet/dwc-qa-manage/issues
Review DCH 7 - Aggregators GBIF and iDigBio

use mics when open to encourage others to do so
use color in the chat!

Old business - that needs followup from last meeting:

- Laura Brenskelle? - presentation - zooarchaeological work (two parts, one prob in common with paleo - trying to accommodate deep time - work with paleo person to put this together)
- TDWG (not DwC Hour) [Walter Berendsohn (TDWG origins were around controlled vocabularies)... historical perspective.] ABCD development. TDWG Steve Baskauf
- How are you using DwC, What you are challenged by using DwC Core? ... 5 - 7 minute talks?
- Latin America outreach region (Colombia, Ecuador, Argentina, Brazil - Antonio Saraiva) CONABIO, Renato De Giovanni (CRIA, Tapir)

When can VertNet present? Get on calendar, book seminar room

29 August 2 PM EDT

Paula Zermoglio host / moderate
sound check Mon 28th - 2 PM EDT
write abstract - DONE and POSTED TO iDigBio

Results of WhenIsGood Poll to review GitHub tickets followup - DONE

all set to go (31/08, 11AM EDT).

DCH 8 sound check schedule

Advertizing TDWG DQIG DCH - any special places, new communities?

DCH 9 Kurator date / time set

around Oct 10th (after TDWG), (Not 3 - 4 EDT - Arctos Webinar at that time)
John needs to finalize TDWG schedule before he can commit to this date / time

DCH 10 Audubon Core date / time set (Gary Motz and Doug Boyer)

DCH 11 (December) - DCH the 1st year-in-review (or a topic instead, if we have a volunteer). Could be a mostly open hour too - to reduce preparation time - encourage microphone use!

ABCD

Walter Berendsohn, Jörg Holetschek

tracking grant / funding numbers - per record. Is this supported in ABCD?

Collection Object Storage Location brought up by Wouter Addink in DwCHour on Aggregators to be able to plan for digitization. Does ABCD support storage location information?

Darwin Core Hour and Q&A management workflow diagram:
https://docs.google.com/drawings/d/1vnDyJB5zl7HS1LZd7OFHq6ZGl5iZdhNEK3ytwbFVtlA/edit

For Issue Resolution meeting, the workflow above will be a good reference, along with
https://github.com/VertNet/dwc-qa-manage

Electronic Cultural Atlas Initiative (Digital Humanities) http://www.ecai.org/
Open Context https://opencontext.org/
also - Ben Brumfield

There could be an Darwin Core Hour on how to submit an issue with a suggestion to a Darwin Core revision/change.

NEXT Meeting? Date / Time?
Thanks Dag for joining us! (Thanks for letting me join, and learn your approach)

Darwin Core Hour 29 Aug 2017 (WEBINAR)- Notes

[The Aggregator's Viewpoint - (More Than Vert)Net]

In this Darwin Core Hour we will follow up in the series on aggregator perspectives with a view from VertNet. Though the taxonomic scope of VertNet as a biodiversity data aggregator is focussed on Chordates, VertNet serves a broader biodiversity data mobilization role that has no taxonomic or geographic boundaries. As an Associate Participant in GBIF, VertNet is also involved in a wide variety of other community services, including the development of, promotion of, and training on biodiversity data standards and data quality. In this webinar we will explore the series of services that VertNet provides, such as migration and data quality processes, as well as its unique extracting and searching capabilities, which allow content such as trait data (e.g., body mass and length) to be sought and retrieved. We will discuss VertNet's role in the broader data mobilization framework and how that relates to other biodiversity data sharing initiatives.

Presenter: John Wieczorek
Moderators: Paula Zermoglio, Deborah Paul

PRESENTATION

History of VertNet
VertNet roles
MaNIS+HerpNET+FishNet2+Ornis = VertNet
1. VertNet - synergies - connections
    - siblings. GBIF - “early parts of VertNet would contribute to, be a part of GBIF”
    with development of GBIF, VertNet was part of early development of the IPT
    - data management systems. Arctos & Specify data hosted by VertNet.
    - Publication synergies:
        - VertNet helped mobilize (host) eBird data at the beginning
        - TEAM
        - PaleoDB, Neotoma, Zooarch
            some of these are converting from old DiGiR format to DwC Archive data sharing format.
    - Development synergies:
        Rebioma, Dimensions of Amazonian Biodiversity, DIPnet
        BioGeomancer, GADM, GEOLocate

    - Consumer services. AmphibiaWeb
    - Support. Provide support: UK, Cyverse, iDigBio,
infrastructure/funding.

2. VertNet Products
    - Publication tools:

- VertNet Migrator Toolkit: Data → Scripts, improve DQ of the data, add additional info (eg, license) → DwC-A.
- Kurator Project (Harvard, NSF). Use VN Migrator capabilities to build actors, which in turn build workflows.
- Data Portal.
- User Guides & Tools: http://vertnet.org/resources/help.html

3. VertNet Community Involvement
- Standards. Participation in TDWG's IGs. Relationship between standards and biocollections communities. Participation in DwC-Hour.
- Training
- Research, published and ongoing.

4. VertNet "How we roll"
- Data publisher discovery. VertNet does not look for data providers, data providers contact VertNet.
- Data publishing guidance. To clean or not to clean data, which license to choose, registration with GBIF, to host or not to host own IPT (VertNet can host). Publication of any taxon as long as provider publishes verts: if already publishing verts, then let's publish it all. (in the VertNet portal only verts, in the VN IPT, all data).
- Data preparation. Use of VertNet Migrator Toolkit.
- Vocabulary management. Lookups from the migrator, vocabularies accumulated to run new migrators.
- Data quality reporting
- Data harvesting
- Data indexing
- Snapshots
- Feedback

5. Biggest Challenges.
- Vocabulary maintenance
- Funding

6. Weak Points. Innovation-based funding model by funding agencies, VertNet: mature products & services that work + capacity to innovate, but needs to support the foundation.

7. Aggregator Future. VN is not going away. Working on finding a sustainable solution.

8. Strong Points:

- Agility for innovation (e.g., traits and tissue searches)
- Staff experience and dedication (est. 1997)
- Scalable infrastructure maintenance costs
  - <$150 per month
- Human maintenance effort required
  - roughly 2.5 people
  - ~80% of effort is for data publisher support

9. Value for Stakeholders:

- Associate GBIF Participant (functionally a node)
- Beginning to end data publishing support, and beyond
- Long-term commitment
- Community involvement
- More than Vert(Net)

10. Value for VertNet: acknowledgements.

COMMENTS

1. KURATOR
Deb Paul: About Kurator: http://kurator.acis.ufl.edu/kurator-web/about. Kurator provides scientific workflow tools for data quality improvement of natural history collections and other biodiversity data. Kurator Web is a set of a user friendly web interface to configure and launch curation workflows while maintaining provenance. Kurator-Akka and the Kurator YesWorkflow data curation software and code are available on GitHub. For more information about Kurator, please visit our wiki. http://wiki.datakurator.org/wiki/

Deb Paul: What data formats (csv? .ods? .xls? xml? dwc-Archive?) can a person use with Kurator?

JW (speaking): you can use any dialect of text, or a DWC-A data format

Joanna (iDigBio): John W: what is the delivery date for first version of new Kurator?

JW (speaking): Kurator is already available and functional, but will be gaining major functionality over the next year or so, and onward.

[DwC Hour on Kurator happening in October 2017!]

2. VERTNET WAITING LIST
Brad Millen, ROM, Toronto, Canada: How many "waiting"?

John W Speaking: Dave Bloom to look for that number.

David Bloom: 1,000,000

rob: LIAR

David Bloom: checking....

Brad Millen, ROM, Toronto, Canada: <G>

David Bloom: A quick review of our wait list is approximately 75 institutions with about 3-4x as many collections waiting for us to get to them - there are at least 12 institution in process with at least 30 collections that are in the midst of the process.

Tom Trombone (AMNH): wow

David Bloom: so, close to 1M

Brad Millen, ROM, Toronto, Canada: That is impressive!

David Bloom: That does not include the institutions who have expressed interest from a Zooarchaeological perspective

3. SCOPE OF DATA FOR VERTNET
cindy opitz, university of iowa mnh: so VertNet can also process invert data?

David Bloom: Yes, plants, inverts, fossils, verts - all biodiversity data
but to differing levels

Brad Millen, ROM, Toronto, Canada: Every time I search on a Genus and species I will get something slightly different then. Like a Box of Chocolates!!

cindy opitz, university of iowa mnh: different levels?

David Bloom: Ha! We aim to make research sweet.

David Bloom: We know more about vertebrates taxonomically than we do inverts so our resolution there is not as robust, but the more we do, the better we get.

Deb Paul: @David B - clarify - you publish the data (so it can go into @iDigBio @GBIF), but only Vertebrate data shows up on VertNet website?

David Bloom: example

David Bloom: Deb - yes, only verts in the VertNet portal, but all biodiversity to GBIF and iDigBio

Deb Paul: @David B - thanks for clarification :-)

cindy opitz, university of iowa mnh: can I send you my insects? when I asked iDigBio, I received a bunch of stuff I didn't understand...your processing of my vertebrate data was so simple!

David Bloom: Yes, Cindy, send 'em. Just put them in Dropbox and I'll put them through their paces.

cindy opitz, university of iowa mnh: :)

David Bloom: I look forward to it.

Brad Millen, ROM, Toronto, Canada: Any file sharing or just DropBox

David Bloom: We use Dropbox for a variety of reasons, but we'll set up the shared folder to avoid any confusion. So, if any wants to share data with us and they do not already have a shared Dropbox that we've setup for them, folks should reach out to me to get started at dbloom@vertnet.org

Brad Millen, ROM, Toronto, Canada: I have Google and OneDrive.

Brad Millen, ROM, Toronto, Canada: Should work nonetheless.

David Bloom: Should. We can talk more offline.

Brad Millen, ROM, Toronto, Canada: Great!

4. MIGRATORS
Gary Motz (Indiana Univ): Can collection managers use the VertNet toolkit themselves to clean and high-grade collection data, especially before it goes back to the collection management software or DwC-A?

Gary Motz (Indiana Univ): Or should we just add to your growing waitlist? ;)

David Bloom: Not yet, Gary, but that is part of the purpose of Kurator.

Brad Millen, ROM, Toronto, Canada: Waiting for Kurator to do just that!

David Bloom: In the meantime, feel free to contact me about getting the job done until you can do it yourself.

JW (speaking): Some discussion of handholding new users trying to learn the migrator toolkit. Some of this documentation will happen during an upcoming workshop in Colombia. Within a year or so the migrator toolkit should be totally phased out by Kurator. Kurator already has a web interface, but has nowhere near the capabilities of the migrator toolkit.

5. HOW DOES VN GET DATA FROM OTHER IPTs
Brad Millen, ROM, Toronto, Canada: How does VertNet get data from other IPTs aside from VertNet`s

David Bloom: We harvest data from IPTs, such as the Field Museum's, in the same way that iDigBio and GBIF do. We harvest the Darwin Core Archive and then aggregate the occurrence records and metadata into our index.

David Bloom: In fact, we harvest from the VertNet IPT in the same way, too.

Brad Millen, ROM, Toronto, Canada: So sometimes Manual and other times by scripts.

David Bloom: Our harvest is triggered manually, but the aggregation and indexing is scripted and automated.

Brad Millen, ROM, Toronto, Canada: Alright, Thanks Dave!

6. PROPS FROM THE FEDS
Alison Loar: very informative! thank you

David Bloom: Thanks Alison.

Alison Loar: We at NPS are really needing to jump onto the aggregated data bandwagon

David Bloom: We're happy to help get you there, Alison. Send me an email and we can discuss where you are, the shape of your data, and how we might proceed.

Joanna (iDigBio): @Alison You have lots of choices.

David Bloom: Data we help you to publish will go to VertNet, iDigBio, and GBIF.

7.
Paula (speaking): [sorry Paula my Adobe Connect died when you asked your question…]

JW (speaking): [and when John was answering…]

David Bloom: Your database should match how you work on a day to day basis.

David Bloom: Let us worry about getting it into Darwin Core.

David Bloom: And I agree with John 100%.

Joanna (iDigBio): most legacy databases are optimized around the label

8. VERTNET FOR THE PEOPLE

JW (speaking): One question I get a lot is why does VertNet provide so much support to collections? The answer is that we believe data publishing is a social infrastructure problem as much as a technical one.

Teresa Mayfield (UTEP): I agree with that! But we need to be trained...

cindy opitz, university of iowa mnh: because you understand how to do it, and some of us don't!

Gary Motz (Indiana Univ): ....and not contributing to the VertNet backlog! but working to reduce it!

David Bloom: To extent that others can do it themselves, we would welcome it, but as a team of people that do this every day, we can provide this support and allow others to get on with other work.

Tom Philippi NPS: from my perspective, VertNet doing it reduces duplication of effort across the community, and it helps with consistency/interoperability.

cindy opitz, university of iowa mnh: you guys rock!

DCH GitHub Ticket Review Meeting 31 Aug 2017

Attending: Deb Paul, John Wieczorek, Gary Motz, Erica Krimmel, Dag Endresen, Holly Little, Paula Zermoglio

Review DwC-qa-manage

- ~~Add permission for editing manage site for Dag, Holly, Erica. DONE~~
- ~~Add column to the series with the announcement links in first column, change content of presentations column (so that title is not repeated). [assigned PZ] DONE~~
- Agreed to talk to ALA and Canadensys in TDWG
- Issue #13. Divide into individual issues and close after. [assigned DP]
- Issue #9. Capture relevant documentation from Google Code archive in Google Doc. [assigned JW]
- Issue #6. Licensing. Include differences data/media. “Ownership” concept. [assigned PZ, all review].
    - **DE>** A photo linked from a Darwin Core record (itself CC-BY) can be more restrictive
- Issue #2. Page to describe new DwC term creation. Process to follow. [assigned JW, all review].
    - **DP>** Quentin could also contribute possibly
    - Follow up with Andy and Quentin
    - **DE>** A process might include describe the need for a new term -- and then followed by exploring if the term is for Darwin Core or one of the other TDWG standards…
    - **HL>** I am perhaps involved with Gary's audubon core work and am part of the tdwg paleo working group considering extensions there
- Timing in our meetings:
  GM> Holly also asked about "How frequently do we all meet?" I'd like to contribute that response to Paula's notes and ask about the frequency of upcoming issue review meetings.
  The moderator (usually me) meets with the presenters of the upcoming webinar usually 1-3 days prior to the webinar. The whole group doesn't meet then.
  The group DOES meet (typically the Friday immediately following the DwC Hour) WITH the presenters (and interested individuals) to give a quick run down on how the webinar went and to create new issues from the questions submitted during the course of the webinar.
  DP> @Holly this is the first formal-ticket-review we've done
  in addition to meeting the "friday" after a webinar (but this is currently loosely arranged).
  GM> Then we divvy up responsibility for tasks like planning (sometimes during that post-webinar rundown) but more explicitly and extensively RARELY in a meeting like today's

- Convert Notes into Issues. Resolve some of this during meetings post-webinars.

Review DwC-Q&A

- Assignment of issues for each of us to take responsibility.
- Convert Notes into Issues. [assigned Erica] If similar to other issues, add as comment to previous issue.
- Issue #80. Time webinar. Reach out to Laura Brenskelle. “State of the issue”
- Extensions Webinar. Contact Dave Watts.

Darwin Core Hour 5 Sept 2017 (WEBINAR)- Notes

Darwin Core Hour: A bite from the core - testing for data quality

Presenters:
Arthur Chapman, Australian Biodiversity Information Services and
Lee Belbin with Blatant Fabrications Pty Ltd

Arthur:

- Review previous chapters that talked about vocabs and data quality.
- Tasks groups in TDWG under the DQIG.
- TG1. Framework on Data Quality. DQ Profile, DQ Solutions, DQ Reports. (Alan Veiga's PLOS paper).
    - https://www.researchgate.net/publication/294444569_Quality_Data_is_Key_to_Improving_Education
    - https://www.researchgate.net/publication/319110107_Fitness_for_Use_The_BDIQ_aims_for_improved_Stability_and_Consistency
    - https://www.researchgate.net/publication/319115928_Defining_a_Data_Quality_DQ_profile_and_DQ_report_using_a_prototype_of_Nodejs_module_of_the_Fitness_for_Use_Backbone_FFUB
    - A conceptual framework for quality assessment and management of biodiversity data
        - https://www.researchgate.net/publication/318004365_A_conceptual_framework_for_quality_assessment_and_management_of_biodiversity_data
    - DQ Fitness for use: data only has quality when it comes to use. Example of how the same record can be quality-OK or Not OK depending on the use.

Lee:

- TG2. Tests and Assertions. Originally to review tools, services & workflows.
- Suite of core tests at data collector and data aggregator levels. Gathering of test-assertions from different aggregators. ~117 tests.
- Example of a test.
- Principles derived:
    - tests will be based on DwC st; DwC terms are verbatim (can't validate) or bound by vocab or extent (checkable).
    - Criteria for test inclusion: 1) informative, 2) easy to implement, 3) mandatory for amendments, 4) in broad use.
    - Null DwC values will not create assertion unless many are missing.
    - Anticipate non-core domain-specific tests.
- Fields describing the tests. Term (FOR THE TEST) --> Description
- Warnings. Examples. Ambiguous, amended, incomplete, inconsistent, invalid, unlikely.
- Examples.
- Next steps.

COMMENTS

1. DATASET GUIDs
Deb Paul: Aha! so each test will have a GUID?

Arthur Chapman - Australian BIS: Definitely will

2. WHAT'S THE TAKE-AWAY FOR DATA USERS?

Deb Paul, speaking: What actions (if any) do you expect the data provider to take?

Lee Belbin, speaking: The results of these tests should be additional fields added to the aggregator data. This is yet to be finalized, but something like additional columns would make the information garnered from these tests visible and available to filter data at the record-level. Currently you can do this within a single aggregator but the process is not streamlined.

Filter for purpose. Results of the test attached to the record wherever that record goes.

Arthur speaking: working with Kurator for annotation method.

3. EXPANDING DQ TESTS BEYOND THE AGGREGATORS

John W., speaking: I'd like to see these data quality tests available for anyone who wants them, e.g. at the level of data providers. Something like having an OpenRefine plugin. Shameless plug for Kurator.

Deb Paul: +1 John!

Lee Belbin - Blatant Fabrications Pty Ltd: Fully agree John

Arthur, speaking: Also agree. Would be even more cool if we could have these data quality tests available at the point of data collection. Have generic code and generic dataset to test against and to be able to translate into any other db systems.

Joanna McC (iDigBio): it would be nice if the tests could be embedded into Specify, EMu and Symbiota, Arctos, etc.

Erica Krimmel: Agreed, Joanna!

Lee Belbin - Blatant Fabrications Pty Ltd: Full agree. It is up to them, but also to us to promote

Joanna McC (iDigBio): I worked with EMu and Specify to get their GUIDs going, so if I can get a good handle on what this all means, I can continue that work.

4. DATA IMPROVEMENTS/ACTIONS

Deb, speaking: When you are designing these tests, how do you decide what action to take? I.e. should the data be fixed, or a warning given, or…?

Arthur, speaking: Many decisions that should be made by the data user. So we are trying to give them enough warnings or information to make informed, consistent decisions. It's not really up to the aggregator to improve the data; rather, it's up to the data users and/or the data providers.

Lee Belbin, speaking: The issue of data aggregators sending feedback to data providers is a big one. Aggregators should have a responsibility of running these tests and sending results back to data providers at a minimum. In some cases if the data providers are not able to address the DQ issues, then the aggregator may see fit to make resources available to seek expert advice for quality improvements on the provider's behalf. Aggregators as a default however are not responsible for the data.

Deb Paul, speaking: So the way I understand the annotations functioning, how could I as a data provider interact with the results of the data quality test? E.g. can I say "hey I've reviewed the results of this DQ test and I think XYZ"

Lee, speaking: this is the issue of provenance. We must track history of change to the data, including assertions from tests and actions taken based on those tests.

Arthur Chapman, speaking: The annotations are supposed to allow a series of comments on the data, as discussed above by Deb and Lee. These will probably be formatted as W3C annotations.

5. WHO SHOULD USE THESE DQ TESTS?

Paula Zermoglio, Univ. Buenos Aires - VertNet: Very much related to what you are saying Arthur, from a dataset provider point of view, if I have some data I would like to test , which would be the exact path I should

follow? Should I wait for aggregators to perform the tests, or should I try to do it myself somehow? And which are the benefits/hassles I would get each way
Arthur Chapman - Australian BIS: Preferably, you would run the tests yourself before uploading it to GBIF or aggregator. Unfortunately not all data providers will do this and will rely on feedback from the aggregator.
Arthur Chapman, speaking: sandbox for the user within an aggregator (eg ALA).
Deb Paul: @Paula - in either any case get the data set out in the world as fast as possible :-)
Joanna McC (iDigBio): the closer to the source you can run the test, the more likely it will be to make them stick.
Deb Paul: ooh @Arthur - nice :-)
Lee Belbin - Blatant Fabrications Pty Ltd: Yes Joanna. Ideally, the tests should be in data capture software for real-time validation.

Post-webinar meeting 5 Sept 2017

- New issues coming from this webinar --> none to turn into an issue.

- Documentation done --> Make issue in manage repo for everyone to vet, once we are agreed, we close issues and documentation is set (for the moment).

- Archive documentation periodically, versioning.
    - Paula Zermoglio, Univ. Buenos Aires - VertNet: where would we archive the documentation?
    - Gary Motz: Publish with Biodiversity Data Journal? Pensoft? TDWG?
    - Gary Motz: BDJ is a stretch...
    - Erica Krimmel: oof that sounds like a lot of work...
    - Gary Motz: (really not that much more work than we're already doing...but for us academic types, it gets credit and citability)
    - Paula Zermoglio, Univ. Buenos Aires - VertNet: what about some kind of "data paper"
    - Gary Motz: What about SPNHC Collection Forum?
    - John Wieczorek: Does this kind of formal publishing lend itself to versioning?
    - Holly Little: isn't there some kind of TDWG publication in the works with pensoft?
    - Deb Paul: RIO
    - Use GitHub with Version release.
    - For publication: minimum viable product = MVP = target for initial release candidate

- Environment more friendly to peopz *see in the future, take Allan’s site as example.

- **Next DwC hour 24th Oct 11 EDT.** Kurator Data Cleaner Workflows

Darwin Core Hour 24 Oct 2017 (WEBINAR)- Notes

Darwin Core Hour - Kurator Web, Presenter: John Wieczorek

Paula Zermoglio and Dag Endresen moderating.

Presentation:

- What is Data quality
    - past webinars about data quality
    - Chapman, A.D. 2005 Principles of Data Quality.
    https://www.gbif.org/document/80509/principles-of-data-quality

- Why should we care
    Value → use
    We don't know all ways in which data might be used
    Make data as fit as possible for as many uses as possible

- What can we do about it
    Kurator
    - Goal: develop an extensible, open source toolkit for workflows to aid biodiversity research using diverse curation services
    - Example data quality workflow (date)
    - http://wiki.datakurator.org/wiki/
    - Actors connected to run workflows
    - Graphical user interface → Kurator Web (http://kurator.acis.ufl.edu/kurator-web/about)
        - Requires login
    - Workflows
        - CSV / DwC Archive.
        - Workflow status. Allows sharing workflows with other peopz.
        - File Uploads. Allows storage and reuse of source data.
    - CSV Darwinizer
    - CSV File Geography Cleaner

- Concluding remarks
    DQ testing and enhancement accessible for everyone, regardless of the technical skills.
    Future Goals:
        Increase the capabilities of the workflows
        Add to the vocabularies lookup resources
        Extend workflows to and from more users
- DEMO

Comments:

1. PARSING COLLECTOR NAMES

Diana Soteropoulos: Will there be a workflow to parse collector first, middle, and last names from a single string to separate fields for multiple collectors?

Paula Zermoglio, Univ. Buenos Aires - VertNet: Hi Diana, I don't think there is currently a workflow within Kurator that does that, but let's keep the question for John as soon as he finishes the presentation

Deb Paul: @Diana, @Paula, yeppers it will be good to hear John's answer. @Diana, certainly OpenRefine can help you with that task, I believe.

David Lowery: @Diana we have some code from FP-Akka which does this, and more importantly does a lookup against the Harvard list of Botanists, but we haven't incorporated this into Kurator yet.

Erica Krimmel, Chicago Acad. Sci: @david, cool feature to look up against the Harvard list!

Deb Paul: Yes @David @Erica - let's expand this beyond botanists. We need to engage the broader community to make this a reality. On a related topic, David Shorthouse would like to work on a People standard for our community to support sharing - linking - enhancing this type of data

Diana Soteropoulos: It would be great to have a tool with a list of collectors/agents that curators can create for their respective institutions to standardize names within a dataset.

Deb Paul: Is this tool / scripts built in to any current collection management software?

Deb Paul: What about get-together's here in Adobe Connect where people bring sample data and we use the tool to see what happens?

Erica Krimmel: Arctos incorporates some similar data cleaning tools; John could likely speak to that too

David Lowery: Specify-HUH. Data structures from the Harvard botanist list are also present in Symbiota.

Matthew Foltz: When it comes to collector names, who do we know that the A. H. Smith in my collection is equivalent to the A. H. Smith in your collection? Seems like we'd be introducing errors if no other data (lifespan, typical collecting localities, etc) are taken into consideration, thoughts?

Deb Paul: @ Matthew, this is why we need a People standard AND Identifiers (like OrcID) for collectors (and researchers and other "agents")

David Lowery: The FP-Akka workflow uses the botanist list (and a short list of entomologists from SCAN), with dates to look for potentially problematic dates of collection. Nicky Nicholson is also looking at clustering collectors in her dissertation.

Sean Thackurdeen: @David - can you provide a link to Nicky's work ?

David Lowery: We've got several implementations of that. Not all are currently available in Kurator-web.

David Lowery: @Sean, she gave a talk on it at TDWG 2016.

David Lowery: @Sean Symbiota and the Harvard Botanist list both have rdf representations of agents using FOAF.

Sean Thackurdeen: @Deb @David - is there an iDigBio group to work on creating an ontology for collectors in the NHC community?

Deb Paul: @Sean, if @David can't provide you links to @Nicky, let me know and I'll connect you to her

Deb Paul: @Sean - I can put you in touch with David Shorthouse. He just proposed working on the People standard and underlying ontology while we were at TDWG17

Sean Thackurdeen: @ Deb +1 yes, to both. thanks !

John (speaking): Not really yet. DwC has fields for collector and preparator that would be appropriate for this type of data validation

David Lowery: @John - actually we do have code for dealing with agents in Kurator, just in kurator-fp-validation, not linked into Kurator-web as well.

2. ETC
Paula Zermoglio: For details on how Darwinizer works, you can look at:
https://github.com/kurator-org/kurator-validation/wiki/CSV-File-Darwinizer
Paula Zermoglio: for File Geography Cleaner you can see details here:
https://github.com/kurator-org/kurator-validation/wiki/CSV-File-Geography-Cleaner
Paula Zermoglio: The vocabularies used by the Geography Cleaner can be found here:
https://github.com/kurator-org/kurator-validation/tree/master/packages/kurator_dwca/data/vocabularies

3. HOW DO I BRING CLEAN DATA BACK TO MY IN-HOUSE DATABASE?
Erica Clites (UCMP): For getting your cleaned data back into your database, that is a function of whether your database can support that?

Deb Paul: @Erica - a key point - probably one of the very biggest hurdles everyone managing collections data is facing. As we get more feedback - how to effectively and efficiently incorporate that information back into our local collection

John (speaking): Yes, depends on your data management set-up in house. If you are using a database you can update info by tables. If you are using Excel you can compare and update or concatenate fields. Another option for Excel users is bringing the Kurator results into OpenRefine to allow more granular updates. Ideally what you might like to do is invoke the same tools that Kurator is doing in the context of your existing database.

Matthew Foltz: it would be great if kurator workflows could be available directly within symbiota/specify, any chance the code could be integrated into those platforms?

Deb (speaking): Collection managers definitely need help making data quality improvements more automated, especially as we receive more quality improvement suggestions from aggregators.

John (speaking): Agree, but the issue is that everyone is managing data in house with different solutions.

Deb (speaking): Would it make sense for people in the future to write this kind of stuff into grant proposals? I'm thinking that people be doing better thinking about how they will actually repatriate data quality improvements?

John (speaking): Yes I think this is a good idea. In Manis/HerpNet/Ornis all the georeferencing work we did is still not 100% completely repatriated to the holding institutions. The reason for this is that our funding proposal included support for doing the georeferencing but not for repatriation.

David Lowery: @Deb non-trivial problem. We've moved from the FP approach of trying to get data directly back into providing data quality reports that users can edit their own data from.

4. CHECK LAT/LONG
Dima Mozzherin: Would Kurator check lat/lon according to geo entity?

John (speaking): Yes, Kurator will be able to check that the provided lat/lon fall within the stated geo entity

Dima Mozzherin: @John, at least that it checks - + for lon

David Lowery: @John and we also have code for evaluating georeferences against textual locality data in kurator-fp-validation.

5. KURATOR ARCHITECTURE?
Dima Mozzherin: @David im curious if you are using thrift to connect to R or other languages?

David Lowery: @Dima We are currently using JNI with embedded Python/R, but looking at thift now and it looks interesting

Dima Mozzherin: @David +1

Dima Mozzherin: @David grpc is another interesting alternative

David Lowery: @Dima Thanks! Have you used either? Would be interested to see how you're using it if you are

Dima Mozzherin: @David we use thrift for https://github.com/GlobalNamesArchitecture/gnindex

Dima Mozzherin: @David we also thinking about plugins based on grpc for BHL indexing

6. RECORD LIMITS
Tomer Gueta: Where can I find the storage size limitation for a CSV / DwC-A file?In addition, are there recommendations for the maximal number of records for a specific workflow?

John (speaking): David might have more input. For DwC there is a limitation on what Kurator can handle, I think about 2GB. So not a number of records but columns x rows, plus richness of data. We don't actually know the full extent of the limitations in Kurator Web yet. Only benchmark so far has been trying to run the full MVZ vertebrate dataset of 1.3 mil records and we ran out of memory problems (this was on a local copy of Kurator, nto web app).

David Lowery: Full MCZ data set, about 2 million records, runs currently on some workflows but not all.

Post-webinar meeting 24 Oct 2017

- schedule meeting for issues review
- Nov webinar: Gary? Deb sent email
- December - review
    - PZ: I would like to see a list of things that we have NOT accomplished
    - HL: Would be interesting to look at the topics that most commonly came up across all of the webinars (from the notes and questions)
    - DP: organize by tags
    - HL: not necessarily count, but really obvious topics
- Extensions DwCHour - include in the list as additional DwCH.

Kevin is making the offline version of dwch8 - a bite from the core - to mp4.

2018 ideas.

Darwin Core Hour 21 Nov 2017 (WEBINAR)- Notes

Darwin Core Hour: Audubon Core and 3D Biodiversity Data: Metadata, Practice, and Unification of Efforts

Date: Tuesday 21 November 2017
Time: 3pm EST, 12pm PST, 5pm ART, 9pm CET, 20:00 UTC
Where: http://idigbio.adobeconnect.com/room

Abstract: A growing mode of digitization of natural history collection objects is 3D digitization, which includes three main acquisition techniques: surface scanning (structured light or laser scanners), volumetric scanning (microCT or MRI), and photogrammetry (structure from motion). There is now burgeoning interest in and tremendous need for describing 3D data files with standard vocabularies in the interest of promoting broad accessibility and long-term digital preservation. Audubon Core is an existing vocabulary and extension to DarwinCore that is used to describe digital media files representing natural history objects. It is not an entirely new vocabulary with many terms borrowed from Dublin Core, Darwin Core and more. It also intended to describe different kinds of digital data representing different creation methods and file formats. We overview several different 3D data collection modalities specifying the details needed for understanding how 3D data was generated and processed. We investigate the utility of Audubon Core for describing these 3D modalities. Questions we ask are which existing terms can be used for describing new 3D modalities, whether new terms are needed, whether certain 3D modalities need specific terms not applicable to other modalities, what 3D data formats should be emphasized for preservation and access, and how to pursue formally acquiring new terms either through creation of new vocabularies or extending existing ones.

Presenters
Gary Motz - Chief Information Officer and Assistant Director for Information Services, Indiana Geological and Water Survey
Doug Boyer - Assistant Professor, Department of Evolutionary Anthropology, Duke University; MorphoSource

Moderator - Holly Little, Informatics Manager, Paleobiology Collections, Smithsonian

Relevant links
https://terms.tdwg.org/wiki/Audubon_Core_Term_List
http://morphosource.org/

From Deb:

GOALS:

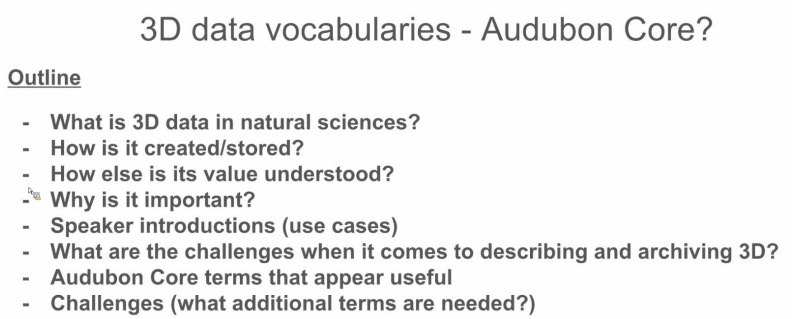
begin collecting known use cases where there are no Audubon Core terms for the concepts and data to be shared.

note terms / concepts we already know need adding

infrastructure issues?
people issues?
mapping issues?

Vocabulary Development - via BIS TDWG (Steve Baskauf). ready to write a charter to tackle this.

Presentation
Overview of 3D data in natural sciences
How is it created/stored
How else is its value understood
Why is it important
Speaker introductions (use cases)
What are the challenges when it comes to describing and archiving 3D
Audubon Core terms that appear useful
Challenges (what additional terms are needed?)



- What is 3D data: imagery that provides 3d representation of object.
- Creation/storing of 3D data: Modalities:
    - Volume based (internal & external information)
    - Surface based (line of sight)
- 2 data types:
    - raster image series
    - mesh / point files

Pointing out one to many relationships and metadata files that need to account for points, shapes, color, relationships, depending on data type.

- Equipment parameters. Somewhat specific to the method (e.g, CT data, Photogrammetry, Laser).
- Some standards not fully developed, people not knowing what they should be reporting.

- Software parameters. Processing required, independent of the method used. Need for comparable parameters. Description of software used for different actions in:
    - CT data (e.g., reconstruction, segmentation, mesh-file creation)
    - photogrammetry (e.g., mesh-file creation)
    - Laser (e.g., segment alignment, mesh-file creation, mesh refinement)
- Associated information on physical specimen. Audubon Core - Darwin Core.
- Visualization. Web previewing. Assessment of data utility by users based on imagery available.
- Post-processing and intended use. Broader uses (e.g., gamification projects, augmented reality, morphometrics). Need for metadata describing original intended use.
- Use case: MorphoSource.org
    - background
- Use case: IU Center for Biological Research Collections. Emphasis on surface scan data. In-house standards. Trying to identify best practices and how to put it together for everyone.
- Challenges for describing 3D data.
    - Best practice standards.
    - describe large variety of protocols, diverse data types, diverse software, etc.
    - documenting connections between data spawned by a single scanning event.
- TDWG & Audubon Core. See: https://terms.tdwg.org/wiki/Audubon_Core_Term_List
- Useful terms in Audubon Core.
    - issues when dealing with modality-specific metadata
    - example: dc:type
- What is missing
    - limited ability to describe 3D data collection parameters
    - examples (modality, #images, energy settings, file size, etc). Some covered by Audubon Core, some others not. Some additional terms needed. Need to determine how such terms would serve different modalities.
    - participation & feedback.

Comments

1. Needs in the collections to incorporate / deal with 3D data

Debbie Paul (iDigBio): How will the creation of all these 3D images affect data management in collections databases (Specify, Symbiota, ARCTOS, etc). What changes might be needed (if any?) to collection management systems to store 3D related data about museum specimens?

Smithsonian Group: @DebbiePaul: in addition to the increased capacity required, the main stumbling block is the long term care and maintenance of these fragile and unique file types. Currently no FITS/JHOVE equivalent (Rebecca Snyder @ NMNH)

Debbie Paul (iDigBio): @Smithsonian Group: where are folks storing the related metadata - about the 3D image files available for a given specimen?

Smithsonian Group: Not consistent. We're working on it. Holly and others are starting to reference the existence of 3D scans in the prep grid of the catalog record.
In future, the idea would be a connection between our CIS (EMu) and the 3D/CT proposed repository

John Wieczorek: @debpaul In Arctos, the database schema accommodates media generically, and links out to the media and the tools to display them, putting that responsibility outside the database. Examples: http://arctos.database.museum/guid/MVZ:Mamm:52125 links out to http://www.digimorph.org/specimens/Taxidea_taxus/.

2. Purpose / use of 3D data (related to discussion 1.)

Debbie Paul (iDigBio): So @Smithsonian Group, is this some of the "fragility" that you are referring to? what might happen to the images after post processing may change or limit their suitability for a given purpose?

Smithsonian Group: I'm thinking the long term management of the bitstream, rather than the use. How do you review, monitor and track bitstream viability other than only hash checks?
Smithsonian Group: How do you know X dicom file or SLT is usable, renderable or corrupted?
Smithsonian Group: Detailed specs on the model will assist in the determination of potential suitability, which is also an important component for providing access

3. Resource Links/Related Projects

Debbie Paul (iDigBio): Morphosource: http://morphosource.org/

Debbie Paul (iDigBio): More about oVert: https://www.idigbio.org/wiki/index.php/OVert:_Open_Exploration_of_Vertebrate_Diversity_in_3D

4. Connections

Doug Boyer: paper proposing standards - http://rspb.royalsocietypublishing.org/content/284/1852/20170194

Doug Boyer: Proposed Metadata Terms for Audubon Core: http://bit.ly/2zkuCXp

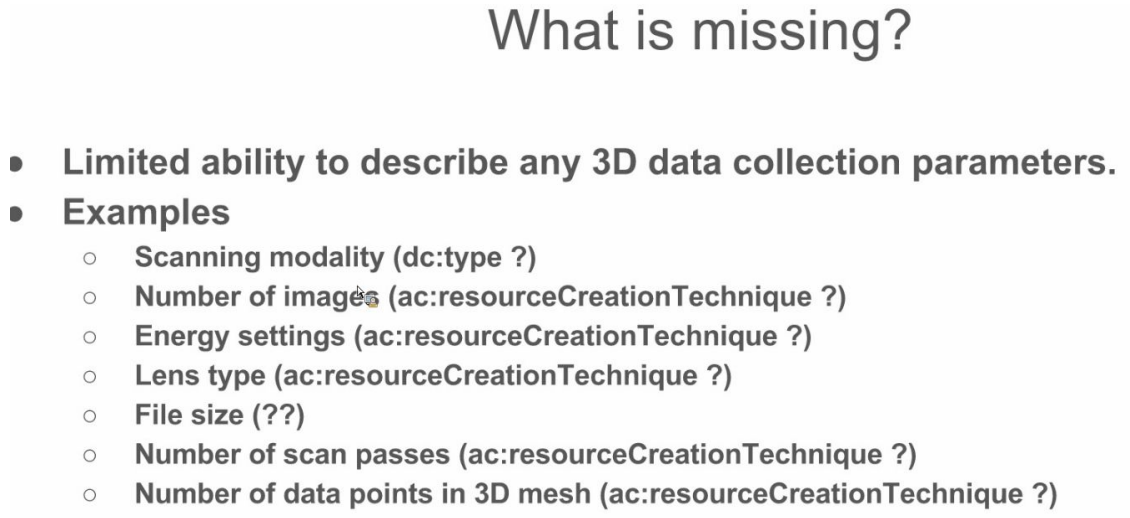
Main Challenges for Describing 3D Data

- **Establishing best practice standards for preservation that are economical**
  - **We'd like to do it right the first time, 3D data is EXPENSIVE to generate & preserve!**
  - **Reproducibility for scientific method integrity ← critical for geometric morphometrics!**
- **Effectively describing a large variety of data collection protocols stemming from a corresponding diversity of equipment and software types (ie finding a balance between creating new detailed parameters and citing workflows).**
- **Documenting connections between all data spawned by a single scanning event**

How do we document the connections?

How do we insure we connect 3D resources being created (to avoid new data silos)?

5. Extending Audubon Core



Proposed Metadata Terms for Audubon Core: http://bit.ly/2zkuCXp

Field Museum: Why not use DICOM vocab for describing volumetric data set?
Field Museum: Extension to Audubon core, rather than modification of AdC?

Holly Little (SI NMNH): The ultimate goal would definitely be to extend AC to include more terms from other standards for sure

Doug Boyer: I think Dicom vocabulary can be adopted for many terms

Holly Little (SI NMNH): especially since AC is already pulling from a variety of namespaces

Smithsonian Group: agreed and there are accepted sharing protocols that use AC and don't accept dicom fields (currently)

BobGhost: ISO has standards for 3D additive manufacturing ("3d printing"). Have these been examined?

Debbie Paul (iDigBio): Do we need a symposium / workshop at TDWG-SPNHC on this?
Stan: @DebPaul: this would be a great subject for a (or several) TDWG/SPNHC symposia.

Laura Vietti -UWyo: Thanks Gary, Doug, and organizers! Would love to be included, or kept in the loop regarding this conversation.. especially now that i'm working to edit our database to include info about our digitized files (scan type, lenses, file format, etc...)

Steve Baskauf: The TDWG Vocabulary Maintenance Specification describes not only how to add new terms to the "bag of terms", but also a process for developing "enhancements" that prescribe more specifically how terms should be used, how things should be linked, etc. See section 4

Stan: Agreed, 3-D data would be a great test for putting together an "application profile"

Debbie Paul (iDigBio): We need examples - of Audubon Core - in use

Gary Motz (Indiana Univ): ABSOLUTELY!

Debbie Paul (iDigBio): and all extensions - for that matter

Stan: This looks like a great common use case where every institution will need to create interoperability between several systems; collection management, DAM
Stan: We will also need new standard for harvesting descriptions of 3D data files

Post-webinar meeting 21 Nov 2017

- **Next DwC Hour: Dec 4th, 11am**. Darwin Core Hour Brainstorming - Community Planning next year's approach. [sound check same day at 10am]. A couple of slides to introduce / guide discussion.
    - Invite some guests.
    - Cross-fertil BITC.

- Add a BITC webinar

- DwC Hour SOMETHING at TDWG/SPNHC 2018. Symposium? “Community building”. Unconference session? Open conversation with burning topics.

Darwin Core Hour 4 Dec 2017 (WEBINAR)- Notes

Darwin Core Hour Brainstorming – Inviting the Community to Plan for Next Year

Date: Monday 4 December 2017
Time: 11AM EST, 4PM GMT (UTC), 1PM ART, 2PM PST
Where: http://idigbio.adobeconnect.com/room

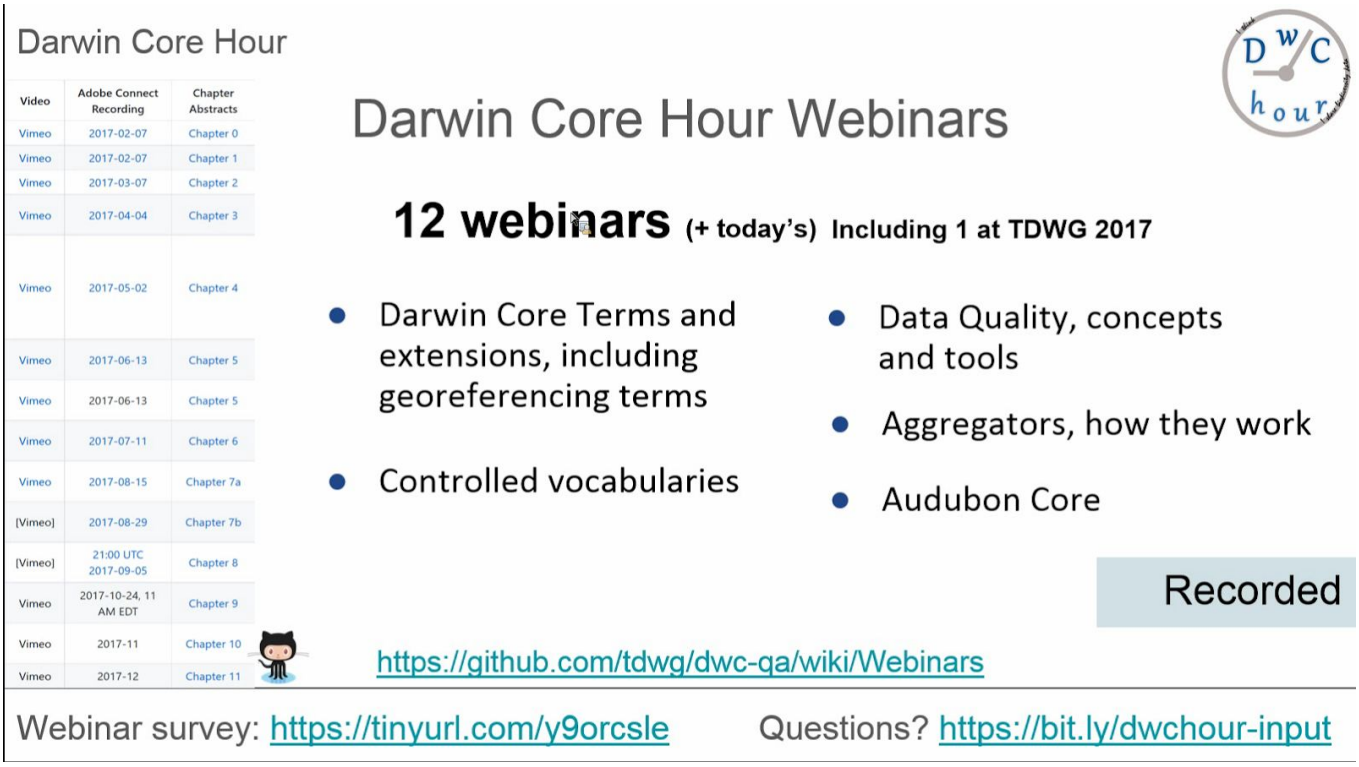
Abstract: Hi there! Our last Darwin Core Hour of the year is here! We would like to take this opportunity to invite you all to an open conversation. During this webinar we will briefly go through the experience of putting together a Darwin Core Hour, i.e., how it works: from the inside. We will assess the topics covered and the ones yet to come, and we will put together a plan for next year. We would love to have you participate, bring your input and ideas for making this initiative grow and address the interests and concerns of the community.

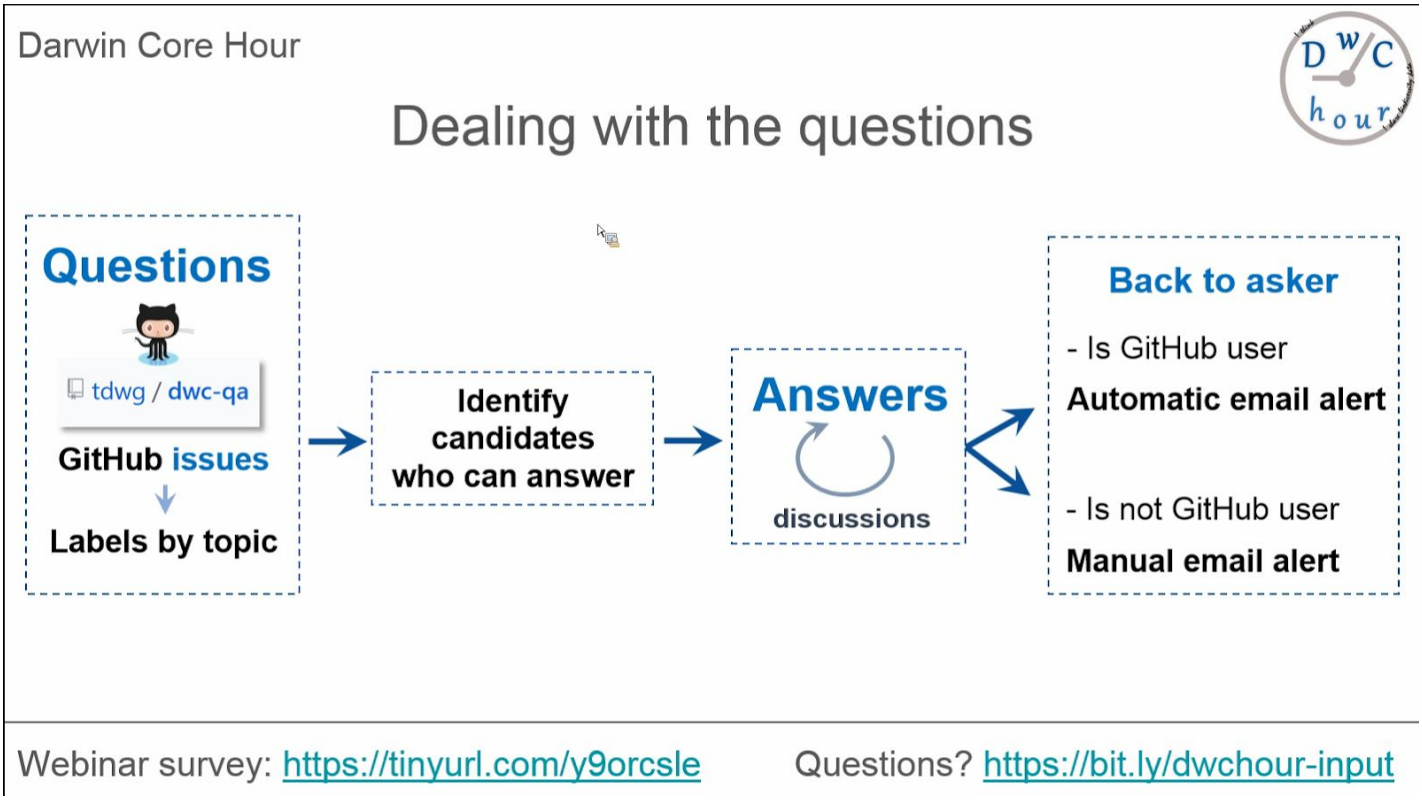
Presenters
Paula Zermoglio
Town Peterson

Moderator - Deb Paul and John Wieczorek

Relevant Links:
DwC Questions: http://bit.ly/dwchour-input
Darwin Core Questions and Answers to them: Synthesized here: https://github.com/tdwg/dwc-qa/wiki
Webinars linked on the DwC Q & A site: https://github.com/tdwg/dwc-qa/wiki/Webinars
Form to submit a question / a request: http://bit.ly/dwchour-input
Your Questions become gitHub tickets: https://github.com/tdwg/dwc-qa/issues
Biodiversity Informatics Training Curriculum (BITC): http://biodiversity-informatics-training.org/training-courses/
Biodiversity Informatics journal:https://journals.ku.edu/index.php/jbi
BITC Website: http://biodiversity-informatics-training.org/
BITC Facebook: http://www.facebook.com/groups/BiodiversityInformatics/

12 Webinars (+ today) in 2017

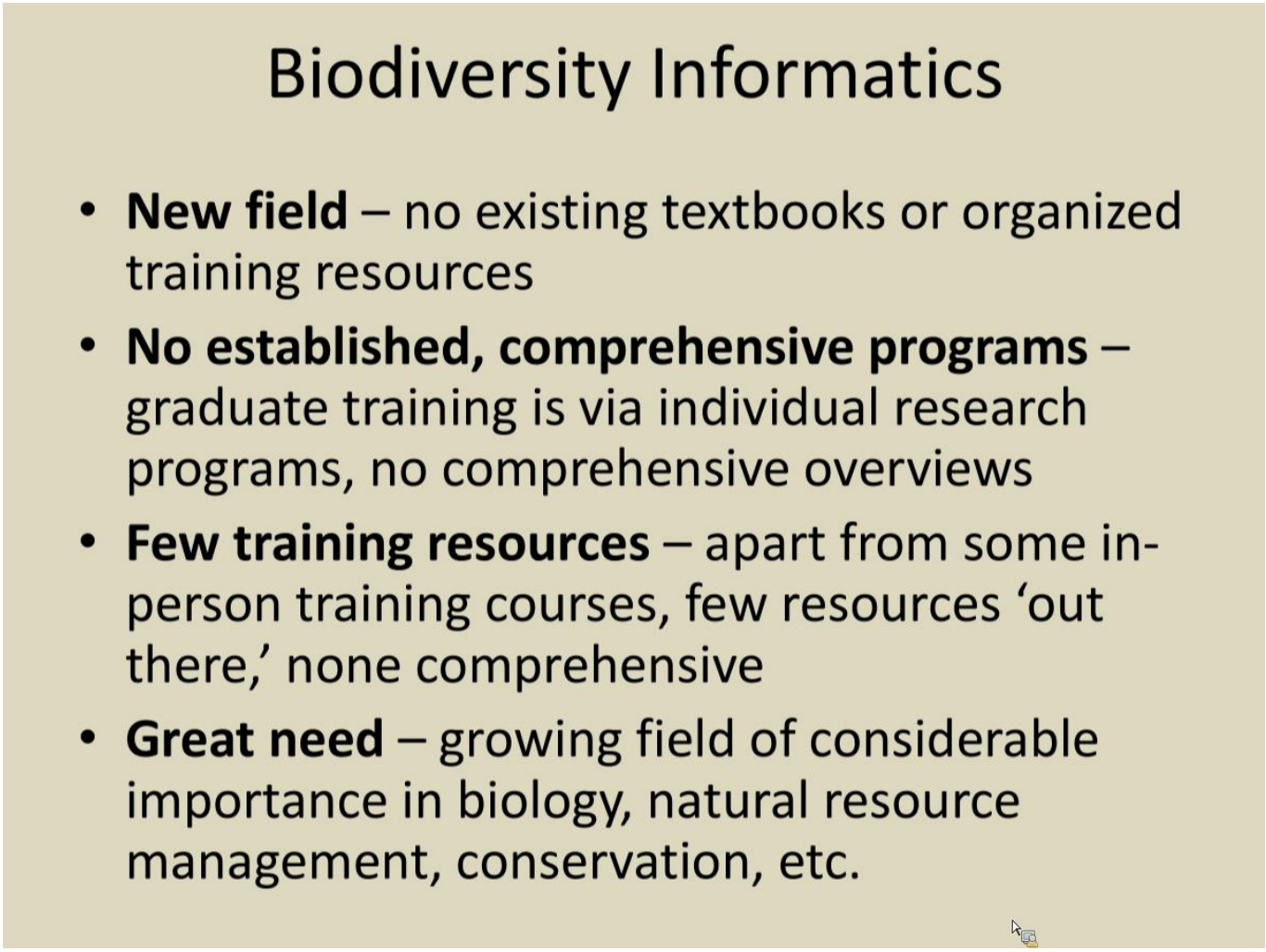




John Wieczorek: The documentation produced on the Darwin Core Q&A site is intended to be auxiliary documentation to the Darwin Core standard - to add examples, experiences, discussions that were difficult to manage openly within the standard itself.

Biodiversity Informatics Training Curriculum, Town Peterson
Digital video based textbook for the broader field of biodiversity informatics. Everything from data capture/data creation to data analysis

Deb Paul: Biodiversity Informatics Training Curriculum (BITC):
http://biodiversity-informatics-training.org/training-courses/



Carry out in person training courses, capture digital video, publish immediately to YouTube, transcribe for subtitles, translate into multiple languages

Resources include open curriculum, active facebook group for members, online seminar series, YouTube channel, open source publication

John Wieczorek: @Town. Paula is new finishing an OpenRefine Tutorial that was proven in courses in English and Spanish.

Town: OpenRefine tutorial???? WONDERFUL! Can we include it in the BITC????

Town: Here are the links for the BITC presentation:

Town: Biodiversity Informatics journal: https://journals.ku.edu/index.php/jbi
Biodiversity Informatics Training Curriculum: http://www.facebook.com/groups/BiodiversityInformatics/ BITC
Website: http://biodiversity-informatics-training.org/

FUTURE PLANS:

- Obtain funding
- Subtitling
- Update courses periodically
- New courses

- Enrich courses with exercises and test data
- Extension to other fields
- In-person courses in other languages
- Cooperative network of graduate programs

Comments:

1. Extending Communications
Dima Mozzherin: may be gitter.im as well for conversations? And apply to stack overflow for biodiversity Q/A?

Deb Paul: Hm. Hi @Dima, I was thinking of a slack channel...

Dima Mozzherin: slack would work too @Deb, gitter.im is more integrated with github though

Dima Mozzherin: https://meta.stackexchange.com/questions/76974/how-can-i-propose-a-new-site

2. Lessons Learned
Town: sometimes the interactive/live events are less accessible to certain parts of the world
Synthesise modules already done for those who cannot attend, compile into course module

3. Including DwC into BITC
John Wieczorek: So, one question for Town. Based on what you have seen of Darwin Core Hours given, do you see them as apt to be listed/included as part of the BIT Curriculum?

Town: Yes.

Paula: This would be great. We may need to have the DwC materials more modular in order to incorporate effectively into BITC.

4. More lessons learned - how to measure effect?
Deb Paul: How to manage / keep up with gitHub tickets: https://github.com/tdwg/dwc-qa/issues

Paula: [overview of DwC team workflow]. Question on when to publish (in whatever form) documentation generated by DwC Hour?

Town: My preference is to get information *out* and then see what happens with that.

John: If you publish content in that nebulus/organic format, do you have a way to see what *happens* to the stuff you release? E.g. versus having a cite-able resource.

Town: From the BITC perspective, I can see who/how many people are viewing resources (via web traffic stats), and I receive comments on a personal level. But in terms of overall effect, it's hard to measure. BITC is testing out a new course format where every Monday we will publish about an hour/ hour and a half of video, and we have labs interested in subscribing to this course. Participants are expected to submit questions and then by a certain day they take the questions and respond to them. This format should help with not depending on logging in for a live event. Enough lag time that people can respond and interact.

5A. Topic Ideas for future DwC Hours: Extensions

Jean W (DMNH): Not related to the BITC discussion, but I'm interested in further discussion on the topic of controlled vocabularies. Second interest is more on decoding the error feedback from places like GBIF to improve our database or our IPT protocol.

John Wieczorek: Two (but not too) BIG and wonderful topics for which we have seen lots of interest.

Deb Paul: Yes @Jean, your second topic is one we'd like to do as well at the upcoming SPNHC/TDWG meeting

Jean W (DMNH): Another topic would be more on DwC extensions- for example, the value of watershed data for environmental consultants has come up for us here and it doesn't seem to be in DwC or any extension

John Wieczorek: @Jean W. Yes, do you think it would be a good option to walk through the process of creating an extension start to finish? Or more about that is important in the domain of watershed data?

John: To clarify, I wanted to know if we needed to cover the socio-technical process or the particular subject matter (e.g. watershed data). If the former, then we have a good example with the new chronometric data extension.

David S: @John - GBIF has a list of extensions in progress? Is that public?

John Wieczorek: @David S https://tools.gbif.org/dwca-validator/extensions.do

Dag Endresen: @David @John, perhsp you mean: http://rs.gbif.org/sandbox/extension/

John Wieczorek: There are stable releases above and releases for extensions under development below.

Jean W (DMNH): We know how we can include watershed data in our database but we're not sure how to get it on a data portal since they seem like they are all strictly DwC compliant. So it seems like we need to add it to DwC but we're not sure if that's the case or not.

5B. Topic Ideas for future DwC Hours: Controlled Vocabularies
Deb Paul: @Jean, on controlled vocabularies - have you seen the resource we created after our DwC Hour webinars on this topic? https://github.com/tdwg/dwc-qa/wiki/Controlled-Vocabularies

Jean W (DMNH): @Deb yes, I've seen this but thanks for the reminder.

Paula Zermoglio, Univ. Buenos Aires - VertNet: Jean, re: controlled vocabularies, TDWG Data Quality Interest Group is about to tackle it: https://github.com/tdwg/bdq/tree/master/Vocabularies. I'd be happy to talk more about it if you are interested

Deb Paul: controlled vocabularies are still an issue. Put together a list of controlled vocabularies that are in use/that we could find. Go to DwC hour Q&A site to see the list of controlled vocabularies from these different places

Paula: Tackling this topic within the TDWG Data Quality Interest Group (see above link from Paula)
Task group within the data quality interest group. Goals 1) create a document where we set a standard format for the creation of vocabularies. Even though there are plenty out there everyone is using their own way to

create them. To be able to interoperate among them we need to have guidelines for the controlled vocabularies to comply to. 2) documenting existing vocabularies?

Jean W (DMNH): @Paula- your link isn't working for me :(but sounds interesting

Deb Paul: Hi @Jean https://github.com/tdwg/bdq/tree/master/Vocabularies

Dag Endresen: @Jean try: https://github.com/tdwg/bdq/tree/master/Vocabularies

Deb Paul: it's the pesky dot at the end - thanks @Dag!

Jean W (DMNH): Thanks @Deb and @Dag- should have seen that myself!

Paula Zermoglio, Univ. Buenos Aires - VertNet: Jean, try through here: https://github.com/tdwg/bdq

Paula Zermoglio, Univ. Buenos Aires - VertNet: and there go into the folder Vocabularies

Deb: When you go to your professional meeting, think about organizing a session on the topic of controlled vocabularies to show what the issues are and get the expertise of the domain experts to help improve the data

William Ulate: Paula, wouild you be reviewing also prior TDWG Standards like the Geographic standard in this process?

Paula Zermoglio, Univ. Buenos Aires - VertNet: hi William, we will not be dealing with other tdwg standards, aside from using the specifications to build the standard format for vocabs

William Ulate: @Laura: Ok, could we add https://github.com/tdwg/prior-standards/tree/master/world-geographical-scheme-for-recording-plant-distributions to the ControlledVocabs Resource list?

Paula Zermoglio, Univ. Buenos Aires - VertNet: @William, definitely!

Paula Zermoglio, Univ. Buenos Aires - VertNet: the spreadsheet is open to add

6. Beyond training - advocacy
David S: What about an advocacy module that targets Chairs of eg Biology Departments to encourage hiring of new research staff whose interests are primarily in informatics?

Deb Paul: @DavidS can you help us make that happen?

David S: @Deb - maybe. Best if it came from young profs, however. They'll know what the market is like.

7. How to increase DwC Hour participation?
Paula: Ideas on how to push this forward and make this grow within the community?

John: Number of people participating in the webinars is basically our only way to measure
We don't really know what our impact is and we can learn a lot from the BITC in terms of increasing that

Deb: We can provide counts on the number of times an adobe connect is clicked on and viewed and can see the count on Vimeo a video is watched. Kevin has to go after the data for Adobe Connect

Town: People can watch a video and we can get those counts, but the question is more, what is the impact? We measure impact in our case by papers published that cite the BITC or graduate students that we have managed to connect the student and advisor… can you measure how much a given extension or a given piece of code is adopted after an event? It's intangible, it's hard to quantify, but it's more meaningful to the extent you are able to quantify it?

John: That brings up an interesting example in controlled vocabulary. Over time we should be able to see if there has been an impact in the use of CVs within the community--see if the "wilderness of values" becomes tamed.

David S: Measuing impact is a bit of a red herring, I'd argue. Only important impact is whether or not someone has been effectively trained and as a result, has been employed.

Deb Paul: The Wilderness: https://github.com/tdwg/dwc-qa/tree/master/data

Town: Exactly, David ... measuring impact as number of hits is cheap

Town: Rather, aim for real impact ... if they are effectively trained, then they will adopt and use the methods or the ideas

Erica Krimmel, Chicago Acad. Sci.: @David, or is currently employed but able to do their job better... continued professional development is pretty essential when the tools and tactics are evolving so quickly in informatics

David S: Start asking for money, for starters. Nominal fee.

David S: Yes, absolutely.starters. Nominal fee.

Dima Mozzherin: Sometimes payments are VERY cumbersome at universities

8. Topic Ideas for future DwC Hours: Data QA/QC
Town: I would throw out another topic or set of topics... how DwC can support effective data cleaning and quality control. That is, can there be a conversation between the users and the data architecture.

Deb: decoding feedback from GBIF, session/symposium at tdwg / spnhc2018.

Jean W (DMNH): @Town- Yes! The watershed data issue came up in talking to a data user. Could data portals provide a pop-up survey to users like commercial sites do?

Jean W (DMNH): @Deb, yes, efficient fixing of errors in some databases is best done through the back end and I expect that many places don't have those resources or knowledge

9. Topic Ideas for future DwC Hours: Usage statistics
John: Third, big topic is one of consistent aggregated usage statistics. E.g. VertNet data providers really take advantage of use statistics provided by VertNet.

David S: @John +1, yes please.

Erica Krimmel, Chicago Acad. Sci.: @John, YES. VertNet stats are great for our board and admin

Deb: Now that we have harmonized the data quality feedback. Some discussion about what would it take to harmonize the usage statistics? Stats from GBIF, ALA, iDigBio and other portals that want to use the same stats parameters
Some of the recent research corroborates researchers learn about data through word of mouth

Jean W (DMNH): @John these stats also show up in grant proposals

Jean W (DMNH): Related to usage stats is tracking papers that have used our data, also key for admin and grants

John Wieczorek: @Jean Another good piece of information to keep in mind for justification.

Dag Endresen: Reusing the same dataset and unit level persistent identifiers can help, https://www.gbif.org/citation-guidelines

David S: @Jean & @John - data citation in literature also means museum collection codes. Tangential to DwC, but perhaps a submodule on data mining.

John: Might be worth a DwC Hour on how to publish to the IPT with DOIs and use that tool effectively.

John: This might be best accomplished with a “state of the state” webinar with a panel, perhaps after the February meeting on data citations. I would be willing to contribute on the metadata side, but would be good to have an expert leading the discussion who works with the citations more closely.

Jim Beach: Some thoughts at KU about how to approach that, sounds like DOI will win out. Maybe there’s a way to create a metadata snippet, probably a DwC extension that would document one usage of one record. So everytime a record was returned as a query result, displayed on webpage, or downloaded… creating some simple plugins that could plug into web servers that would report these snippets, more item level usage metadata more quickly than literature tracking

Deb: at iDigBio you can see the number of times a particular item appeared on someone's screen and then if they viewed it that means they actually clicked on it. What we need is to figure out what levels we need for meaningful stats.

Jim Beach: somebody looks at my record, does that count as a research utilization? Maybe a controlled vocabulary for usage types, sort of codify the usage types into a reasonable number of categories and then create these plugins that would generate those usage snippets of whichever category they were being generated.

Deb: a lot less likely to compare apples to tangerines. For controlled vocabulary discussions like that, a nice topic for showing what is available, eg what does GBIF do what does iDigBio do? This data is documented

Erica Krimmel, Chicago Acad. Sci.: @Jim when you're talking about this plugin, is it something you're envisioning as a feature in Specify or mroe broadly available?

Jim Beach: Probably specify 7 (web based). Looking at how they used Life Mapper?
For most web servers it would be fairly easy to build a standardized plugin that people could stitch into their web apps to generate those things.
Would want some community context for the software solutions

Jean W (DMNH): @Jim- I like it! Important to keep in mind the need to summarize for annual reports. GBIF has detailed usage stats linked to each search but there is no way to summarize them

David S: Gonna require some cultural shifts in taxonomic community because they cite specimens in a way that may not be conducive to us tracking eg DMNH Bird 85448.

Matthew Collins: David S, yes, thank you. Hand constructed abbreviations of catalog numbers are not machine readable.

Jean W (DMNH): @David S I think the bigger problem may be the non-taxonomic community who are probably using large data sets and may not understand the importance of citing the source at the museum level (and the bigger the dataset the bigger the challenge of doing so)

10. Topic Ideas for future DwC Hours: Connecting people to other people
Deb: Another topic, identifying the taxonomic experts from your collection, help exposure, how to share that data and what is should look like in a way that isn't a comma separated list, make it easier to find

11. Collections-level metadata
Deb: When people go to fill out the EML form in the IPT, they may conflate the resources associated with the collection versus the dataset. Couldn't we track metadata as part of our collections databases?

Erica Krimmel, Chicago Acad. Sci.: I've thought about this before and agree with you whole-heartedly, Deb.

Erica Krimmel, Chicago Acad. Sci.: I want dataset-level metadata integrated into our DB so that we can have certain things (e.g. geographic scope or major collectors) automatically generated. Then we can supplement but don't have to waste time tracking down basic stuff, as you described

Jim Beach: Erica, good idea!

Jim: Long story short, we need to make sure that metadata improvements are connected to the stakeholders. How can we make sure that collections impress the local stakeholders who control the purse strings?

Deb: Current initiative focused on re-envisioning how collections-level metadata is shared. Similar to what is available on GRBio but with a level of information similar to that of datasets on GBIF.

------ **DISCUSSION**
To discuss:

- Scope: have we accomplished what we had expected
- Reach: expected vs actual audience

- Effort: is the initiative sustainable

To plan:

- Future webinars
- Pendings
- Outreach & collaboration: synergies
- Management: effectiveness & efficiency

Post Webinar

Paula: How to get more people involved? Bullet points of things to get done with specific delivery dates?
John: In any particular part?
Erica: Suggestion to focus on writing up the documentation
Paula Zermoglio, Univ. Buenos Aires - VertNet: + 1 Erica
Jim: Find people whose survival depends on this kind of metadata… people that need to justify the ongoing support, people that understand the power of that kind of connection, would be more motivated to help
Put out a survey to ask who wants to defend their collections

Regarding redirecting a DwCH question to eg. @tdwg it seems that there is a way to redirect to a team or group: https://github.com/blog/1121-introducing-team-mentions

2018 NOTES

Email sent to DwC Hour Team Jan 31 2018
Hi All,
Welcome to 2018 and February is here already! We need to catch up on DwC Hour. Here are a few items to start off with.

1. **Webinar Ideas** we have documented in the DwC Hour Google Doc Notes. Need more development to move forward. Some will need tickets created (if we decide to pursue).
2.
    1. TDWG (maybe not DwC Hour?) [Walter Berendsohn notes **TDWG origins were around controlled vocabularies**)... an historical perspective presentation would be good] ABCD development. TDWG - Steve Baskauf may be person to ask.
    2. **How are you using DwC**, What you are challenged by using DwC Core? ... 5 - 7 minute talks?
        1. who might we ask - rope in - to do this?
    3. Latin America outreach region (Colombia, Ecuador, Argentina, Brazil - Antonio Saraiva) CONABIO, Renato De Giovanni (CRIA, Tapir)
        1. Any volunteers to reach out to Latin America? What do we want them to talk about?
    4. **Extensions Webinar.** I see a note to "Contact Dave Watts" in reference to an Extensions Webinar.
        1. Do we want to go forward with this? I don't know Dave Watts - so perhaps someone else can do this?
    5. **BITC Webinar.** Not sure who suggested this (John?).

    6. **Update on Vocabularies Webinar?**
    7. **Webinar on the DQ feedback?** (at least some of this is directly relevant to data standards discussion and following darwin core).
        1. Jean W (DMNH) asked for such a thing.

    8. John, you mentioned wanting to strategize for how we might integrate our efforts with BITC? I propose a meeting with Town, et al to discuss how this might work.
    9. Maybe a DwC Hour on - **Preview of DwC Hour relevant content at SPNHC-TDWG?**

3. **Meeting strategy.**
    1. For 2018, do we want to meet 1x month, or continue to meet after scheduled webinars?
4. **Q&A site management. Review what needs doing.**
5.
    1. Do we want a separate meeting (every other month maybe?) to review tickets and progress?
    2. **TODO**: Please review tickets (assign, update, close where you can) before next DwC Hour if possible.
    3.
6. **New members.**
    1. More names of people to invite/include please?
7. To Paula,
    1. I think you wrote this: "Environment more friendly to peopz *see in the future, take Allan's site as example."
    2. Paula can you remember what this is in reference to, and who is Allan? This comment is in our google doc notes for meeting on 5 Sept 2017

3. You also requested we review - what needs to be done (on the QA site) and what webinar status is.

8. **SPNHC-TDWG 2018.** Last year we presented at TDWG. Do we want to present to the SPNHC audience? Lowest hanging fruit - we could submit a request for a Special Interest Group (SIG) meeting - to get interested SPNHC members to talk with us about participation in these efforts. [I think we could get this added - they can be somewhat ad hoc - but are space dependent].
9. Wowsers - you made it to 7? Time for a cup of tea? coffee? something involving yeast?

Of course you will all have items I didn't list - fire away!

Note that' I've also put this list in our Google Doc

With gratitude,
Deb

Darwin Core Hour Notes for 2018

https://docs.google.com/document/d/1rL6KwWSs8SiNmBjkXDeJgOI2QxqIFBE0X9p9NaiSSYM/edit?usp=sharing