Topic: GBIF Unified Data Model for paleo (and geo) specimens
Date: 2023 July 13
Time: 12 PM Eastern

In Attendance

Ben Norton, NCMNS	Erica Krimmel	Janaki Krishna (UMNH)
Roger Burkhalter, SNOMNH	Lindsay Walker, Symbiota	Holly Little, NMNH
Amanda Millhouse, NMNH	Natalia Lopez Carranza, KUMIP	Jessie Nakano, NMNH
Ducky Nguyen, NMNH	Susan Edelstein, NCMNS	Kit Lewers, CU
Jacob Van Veldhuizen, CUMNH	Wai-Yin Kwan, Whirl-i-Gig	Talia Karim, CUMNH
Nicole Volden, NMMNH		

Resources

- Paleo Data Working Group Google Folder
- Paleo Data Working Group website schedule of upcoming Happy Hours and recordings/notes from past Happy Hours
- Tim Robertson at Digital Data 2023 on the new data model: https://youtu.be/CNRAJvyDHu8?t=1268 (start at 21:15 minutes in)
- GBIF blog post about new data model (including links to further reading and webinar recordings): https://www.gbif.org/composition/HjlTr705BctcnaZkcjRJq/gbif-new-data-model
- More in-depth read: Diversifying the GBIF data model
- See an example use case of how a specimen with media looks in the new data model: 07_specimens_with_media
- Materials on GitHub from different CMS: https://github.com/gbif/model-material
    - Instructions for the data mapping exercise: https://github.com/gbif/model-material/blob/master/data-mapping.md
    - Most up-to-date schema: https://github.com/gbif/model-material/blob/master/schema.sql
- 20230713_paleo-happy-hour_slides

Agenda

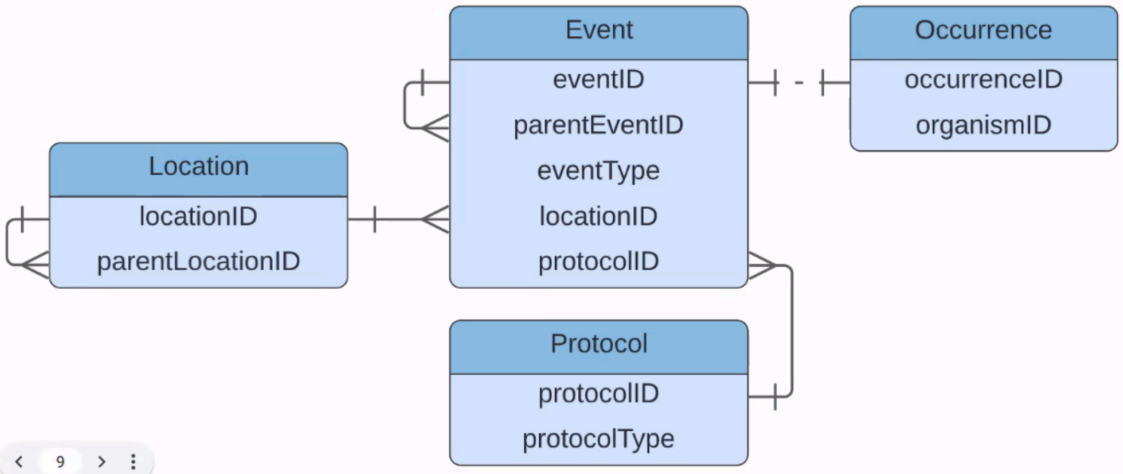
- Announcements/Updates
    a. Upcoming Happy Hour on July 27
    b. No happy hours in August

- 20230713_paleo-happy-hour_slides

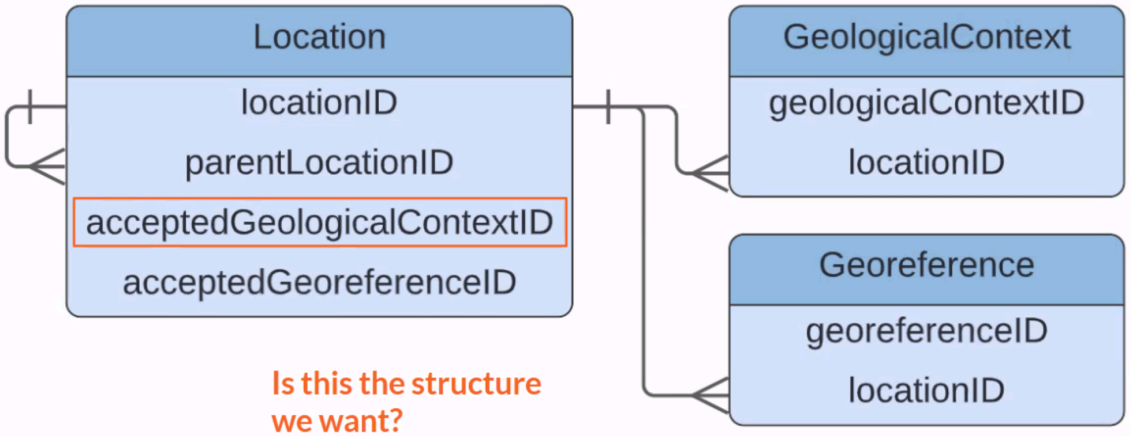
Notes:

- Darwin Core → New Data Model (UM) (increased complexity from DwC)
    - Could this enable us to better share paleo and geo data?
    - UM is less “flat” than DwC
    - Would incorporate more types of data (beyond paleo/geo)
    - Allow more entry points into our data
- Paleo data currently very specimen centric, but the UM is not
    - Core component is the Event + Entities (material, digital) that relate to the Event
    - Current DwC mapping is supposed to translate to the UM
    - Implications for Geological Context, compound specimens (see last HH), multiple IDs within lots, and how we publish our data to the new UW
- UM testing
    - https://github.com/gbif/model-material/blob/master/schema.sql
    - https://github.com/gbif/model-material/blob/master/data-mapping.md
    - NMNH testing for paleo data, but looking for input from others (due end of Aug)
    - Paleo examples
        - Event/locality with multiple specimens of multiple taxa
            - Could include multiple time periods
        - Site with multiple events
        - Site or fossil that was not collected but captured digitally
        - More examples?
            - Cores
            - Measured sections
            - Do field observations count? e.g. I saw 16 inoceramids here
            - Faunal lists
    - Next step for GBIF is to think about how the data publishing models will look
- Things for us to think about
    - Are there things we’ve discussed for DwC that the UM could address?
    - What data normalization will we need?
    - Where are local identifiers (like locality numbers) going to go if we need to use GUIDs to support the UM?
- What will be visible and queryable in the new interface?
- UM can help inform our own internal CMS-es moving forward
    - Arctos is leading in some ways

Event as the core concept



Geological Context linked to Location



Is this the structure we want?