

# **Recommendation of the C4P Working Group: A proposal regarding the paleontology data model used by Specify**

Washington, D.C. – May 21-23, 2014

Gil Nelson, Florida State University (gnelson@bio.fsu.edu)  
Ann Molineux, University of Texas at Austin (annm@austin.utexas.edu)  
Edward Davis, University of Oregon (edavis@uoregon.edu)  
Michael McClennen, University of Wisconsin-Madison (mmcclenn@geology.wisc.edu)

## **1. Introduction**

The following proposal, draws upon discussion of the document “Specify 6's Approach to Stratigraphy: The Specify 6 Paleontological Data Model”, dated 31 March 2009, by the Specify Software Project Staff.

## **2. Proposal**

The central elements of our proposal, outlined below, are designed to preserve existing table structures while offering flexibility to Specify users.

- 1) A separate entry/editing function for Paleo Context records should be created, similar to those for Collecting Events and Localities. This screen should encompass Lithostratigraphy, Chronostratigraphy, and Biostratigraphy. In other words, Paleo Context should become a first-class entity in the database, on the same level as (but separate from) Locality.
- 2) Paleo Context records should be identified by user-specified and user-visible names, in the same manner as is currently done with Collection Event and Locality records. These names would be in addition to the internally generated key field used to uniquely identify these records and link them to other tables.
- 3) The entry/editing screens for Collection Objects, Collecting Events, and Localities should each be modified to allow users to enter a Paleo Context name.
- 4) The Paleo Context name associated with a Locality (if any) should be inherited by all Collecting Events that are linked to that Locality, unless explicitly overridden. The user should be able to override this for individual Collecting Events by specifying a different name in the Collecting Event entry/editing screen.
- 5) The Paleo Context name associated with a Collecting Event (if any) should be inherited by all Collection Objects that are linked to that Collecting Event. The user should be able to override this for individual Collection Objects by specifying a different name in the Collecting Object entry/editing screen.
- 6) The Paleo Context name entered for (or inherited by) each Collection Object should automatically create a linkage between that Collection Object and the correspondingly named Paleo Context record.
- 7) If no Paleo Context name is entered for (or inherited by) a Collection Object record, then it should be linked to a new unnamed Paleo Context record.

Making these changes will allow paleontologists to use Specify efficiently while preserving the current Specify table structure. Because the table structure need not be changed (except for adding a few name fields), we hope that the amount of programming work involved will be small enough to make the proposed changes feasible given the resources available to the Specify Software Project.

### 3. Justification

The fundamental reason behind our proposal that Paleo Contexts be given user-visible names is that these names would serve the same basic purpose as the Collection Event and Locality names currently used in Specify 6. Since Paleo Contexts and Collecting Events are important pieces of data in their own right that each need to be linked to many other records, it seems eminently reasonable to us that users be provided with a mechanism to do this.

In Specify 6, Paleo Contexts are anonymous. There is no way to associate new records with already-entered data; rather, that data must be laboriously typed in again for each new record. Allowing us the ability to name and link Paleo Contexts will make entry of paleontological data much more efficient.

Assuming that you are able to make this change, the question then becomes: should Paleo Context records to be linked to (1) Localities, (2) Collecting Events, or (3) Collection Objects? All three possibilities are discussed in the document "Specify 6's Approach to Stratigraphy", leading to the conclusion that Specify 6 will use approach (3).

Unfortunately, the procedures and data models in use by different institutions vary widely enough that each of these approaches works well for some and badly for others. In fact, in our small working group, we had passionate advocates for and against each one. Some institutions enter separate Localities for each different 3-dimensional location from which objects are collected, leading to a preference for linking Paleo Context to Locality. Other institutions enter a single Locality for each latitude/longitude position, and enter multiple Collection Events each covering a single stratum. These institutions would prefer to link Paleo Context to Collection Event. Still other institutions enter Collection Events that cover multiple strata, and so would prefer to link Paleo Context to Collection Object.

Given that approach (3) is now built in to the Specify 6 table structure, we recognize that this table structure is infeasible to change. We understand that this is due both to the programming cost involved and because of the disruption entailed in switching existing Specify 6 installations to some other database structure. That said, we believe that our proposed approach will satisfy the community of Specify 6 users who handle paleontological data while requiring a minimal amount of programming work to implement. **In a nutshell, we propose that the basic linkage be kept between Paleo Context and Collection Object, using the internally generated key field from the Paleo Context table. However, users should be allowed to optionally associate a Paleo Context *name* with each Collection Event and Locality record, and these names should be inherited by all linked Collection Event and Collection Object records unless a different name is specifically entered for those records. Those names would then be used to generate the actual record linkage at the Collection Object level.**

This system would provide the necessary flexibility to satisfy all of the different models discussed above. Institutions whose procedures specify a uniquely named Locality for each 3-d position from which collections are made would be able to create a named Paleo Context corresponding to each Locality, link it to the Locality record by specifying the corresponding name, and then this would automatically propagate to all Collection Events and Collection Objects linked to that location. On the other hand, institutions whose procedures specify a uniquely named Locality for each 2-d geographical position would not enter a Paleo Context name when creating or editing Localities. Instead, they could create a

named Paleo Context corresponding to each unique stratum from which collections are made, and link each Collection Event to the corresponding Locality and Paleo Context. The Paleo Context names would then propagate from the Collection Events to all of the Collection Objects linked to them. Finally, institutions whose procedures specify that Collection Events may cover multiple strata would enter Paleo Context names only when entering Collection Objects.

#### 4. Followup

We would appreciate a response from the Specify Software Project programmers as to the feasibility of this proposal, and would also like the opportunity to answer any questions that might arise in the process of considering it. Please e-mail the members of the working group, as listed at the top of this whitepaper.

